CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Research Notes

Issue 66 • November 2016

The 2015 Cambridge English/English UK Action Research Scheme

A quarterly publication reporting on teaching, learning and assessment

ISSN 1756-509X

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Research Notes

Issue 66/November 2016
A quarterly publication reporting on learning, teaching and assessment

Printed in the United Kingdom by Canon Business Services

Research Notes

Issue 66 • November 2016

Contents

Editorial

Cambridge English continues its involvement in action research for teachers with this issue of *Research Notes*, which presents five papers from the 2015 Cambridge English/English UK Action Research Scheme. This scheme supports English language teachers working in schools and institutions who are members of the national association for ELT, English UK. In the introductory article, our scheme mentor, Simon Borg, explains the contribution of the scheme to the wider teaching community and considers how the challenges facing action researchers can be overcome.

Fiona Wattam opens with an article describing her attempt to improve student progress through the revision of her approach to writing correction. A review of the literature on error correction revealed opinion was divided on the topic, but Wattam decided to employ a correction code without error description to her marking to encourage learners to work out their own errors and investigate whether this would have short-term and/or long-term benefits. A group of pre-intermediate students were selected to trial the correction code, with a post-trial questionnaire revealing a generally positive response to this approach. This inspired Wattam to undertake a second cycle of research where errors were simply underlined then orally corrected by the student. Through the discussion of four individual case studies and questionnaire data, Wattam reveals that the new approach enhanced her capacity to listen to students and expedite the marking process.

Speaking tasks are the focus of the second article, where Lindsay Warwick examines formative assessment, in particular the area of 'success criteria': a set of criteria given to students prior to tasks that outline how to be successful in the task and the ultimate learning objective of completing it. Her research considered whether such presentation of success criteria can improve student self-assessment and speaking skills. Through comparing her own ratings with that of her students, she found that students' accuracy did increase significantly when shown the criteria prior to the speaking task, although students generally did not perceive a difference between receiving the criteria before or after. Overall, Warwick was aware that the encouragement of self-assessment was beneficial to students' understanding of their own progress.

Richard Flynn and Christian Newby continue the theme of self-assessment by examining how to ensure the self-awareness of learners through a study of low-level Middle East students. Through face-to-face interviews and questionnaires on pre-assigned tasks of speaking and writing, which were refined over the course of two cycles of research, Flynn and Newby found that specific criteria and a clear target model encouraged the most accurate self-assessment, and that, by allocating specific time to self-reflection and private discussion with teachers, learners became more willing to consider ways of improving their performance and addressing particular areas of concern. Furthermore, the authors became aware of the need to constantly refine their data collection instruments in order to respond to the learners' limitations as much as to encourage their progress.

Both formative and summative assessment are under investigation in the next article, which is written by April Pugh and Ceri Thomas. As they felt their weekly discrete items tests on grammar and vocabulary did not match the productive language use tests given at the end of term, they decided to research what influence a focus on productive skills would have on students' performance of writing and speaking tasks in the mid-term and end-of-term exams. Through the use of questionnaires and testing one group with productive skills and another with the standard discrete item test, the teachers found that students in the former group responded more positively as they had clearer expectations of assessment criteria, but both groups appreciated the ability to engage in dialogue with their teachers.

Finally, Andrew Taylor reports on a project that also uses a contrast, this time between traditional and online homework. Taylor wanted to challenge his assumption that students would be universally digitally literate and prefer online homework over the paper-based variety. His first research cycle used questionnaires to find out learner preferences and whether they would be open to the idea that the online self-study tool E-Learning would improve their learning. In his second cycle he provided his learners with a homework feedback sheet and assignments in contrasting media, to allow students to express their preference, if indeed they had one. Although students did provide extensive comments on the homework medium, the research inspired Taylor to consider the value of regular student feedback on tasks in order to create tasks that suit individual learner needs, regardless of the medium.

All of the studies presented here show the value of placing the classroom and the learner at the centre of assessment research. They also offer a foundation for future action research, which Cambridge English continues to support and engage with, currently through two schemes in the UK and Australia.

Action research: Not just about 'results'

SIMON BORG BERGEN UNIVERSITY COLLEGE

Introduction

This issue of *Research Notes* presents the the work of the second cohort of teachers on the Cambridge English/ English UK Action Research Scheme. I have described the scheme in more detail in issue 61 of *Research Notes* (Borg 2015:3–5) but in essence it runs for 10 months and includes three face-to-face workshops and online support for the teachers in between these meetings. The overall aim of the scheme is to support teachers in conducting action research projects in their classrooms.

It is very satisfying for everyone associated with the scheme to have these reports available for wider dissemination. Sharing action research in this way is important, for while teachers' inquiries focus in the first instance on developing local understandings of teaching and learning, they should, as with any form of research, seek to contribute to knowledge more publicly. Publishing these reports is one way the scheme makes such a contribution; a second strategy is giving teachers space to talk about their work at the English UK Teachers' Conference in November each year. This for me (and I suspect for the teachers too) has always been a highlight of the scheme.

Here is a brief overview of the projects that are included in this issue (I will not pre-empt reader curiosity by revealing the findings though).

Fiona Wattam's paper is about teaching writing. She examined whether the use of specific corrective feedback strategies by the teacher and asking students to redraft their writing had an impact on students' ability to identify and self-correct errors in their written work. Over two 4-week cycles, 12 students received feedback first via correction codes, then, less directly, through underlining. What Fiona learned from this project challenged her expectations about about what it was that students valued most in the feedback they received from teachers.

Lindsay Warwick examined the use the of assessment criteria in speaking tasks in the classroom. Motivated by more general educational work on formative assessment, especially the idea that prior knowledge of 'success criteria' might enhance performance, Lindsay investigated whether presenting assessment criteria before tasks would allow students to self-assess their performance more effectively; she also looked at whether any improvements in self-assessment were reflected in actual improvements in their speaking skills (as assessed by the teacher). Nine students took part in the project over seven weeks.

Richard Flynn and Christian Newby studied the impact of weekly self-assessment of written tasks on the autonomy of low-level Middle Eastern learners. This project was motivated by the authors' experience that such learners often lacked the skills and dispositions required for autonomous learning and they wanted to see whether making self-assessment a regular feature of their courses might address this issue. Three participants took part in Cycle 1 of the study and another three in Cycle 2, with each phase lasting four weeks.

In the fourth paper, April Pugh and Ceri Thomas took as their starting point what they saw as a mismatch between the productive summative assessments their students had to complete and the discrete item formative assessments that these students were given by way of preparation. In response to this situation, they introduced productive writing and speaking formative assessments and examined the impact these had on students' performance in the summative tests they did. The study unfolded over two 6-week cycles, with 10 students in the first cycle (six intervention and four control) and eight in the second cycle (three intervention and five control).

Finally, Andrew Taylor focused on students' attitudes to conventional and online homework. This project was motivated by the move in Andrew's school to implement an online learning platform which provided learners with a wide range of exercises they could complete for homework. Twelve adult learners took part in the study, over two 5-week cycles. Again, what Andrew discovered about students' attitudes to homework challenged his (and perhaps his school's) assumptions about the ways in which students respond to opportunities for online language learning.

These projects reflect many key characteristics of action research (see Burns 2010):

- the topics studied were chosen by and of direct relevance to the work of the teachers
- teachers' overall concern was improving the educational experience of their learners
- teachers' introduced an innovation into their work and evaluated its impact
- they collected different kinds of evidence to evaluate their projects
- this evidence was analysed systematically
- the projects evolved through cycles of action, reflection, modified action and further reflection
- the work has been made publicly available for fellow professionals to review, learn from and build on.

Additionally, in two cases the projects were collaborative, and while this is not a requirement for action research it is something that the scheme encourages; in practical terms collaboration means the workload is shared but the benefits extend beyond that, and allow evidence to be collected from different classes and to be analysed in greater critical depth.

Continuity and scale in action research

Our experiences of the Action Research Scheme this year have prompted me to reflect on the particular challenges that the UK ELT sector creates for teachers wanting to engage in this form of professional development and I will now comment on these challenges.

In state schools around the world, a teacher knows they will work with the same class of learners for a whole school year and in many university or pre-university contexts language courses are also of a fixed (if shorter) duration (e.g. 10 weeks). In both of these contexts, teachers wanting to conduct action research can plan projects in the knowledge that they will be working with a certain number of learners for a pre-defined period of time. The UK ELT sector, however, is characterised by a large number of EFL schools where rolling enrolment is the norm. What this means is that international students will arrive and leave on a weekly basis and while some students may stay at a school for several months, others will be there for shorter period of time (in addition, it is not unheard of for teachers' timetables and the classes they teach to be changed at short notice). This lack of stability is a major challenge for action researchers in this context because it becomes very difficult to work with a consistent group of students over an extended period of time. This in turn complicates the task of reaching meaningful conclusions about the impact on students that teachers' interventions are having.

If we look at the studies that are presented in this issue of *Research Notes*, we can see how teachers have attempted to respond to such a challenge; they have tried to identify a core group of students who they know will be studying with them for an extended period (typically 4–5 weeks) and to make them the participants of the study; other students may join and leave the class during the period of action research, but they will either not be involved in the study or if they are, any data they provide will not be analysed. A consequence of such attempts to work with a consistent group of learners over a number of weeks is that the groups teachers do eventually work with tend to be very small; Flynn and Newby were only able to involve three students in each of their two cycles while for Pugh and Thomas their two intervention groups were made up of six and three students respectively, with four and five in their control groups. Such sample sizes are problematic when an attempt is being made to use quantitative data to reach meaningful conclusions about the impact of an intervention on a learning outcome and it is an issue that has troubled me for some time. Let me try to articulate here my feelings on this matter and to offer some thoughts on how it might be addressed.

Firstly, it is important to stress that while continuity is an important facet of action research, continuity does not have to mean continuity of *participants*. Continuity can be achieved through an extended period of study having a consistent focus, even though the individuals taking part may vary over time. For example, students' attitudes to particular language learning tasks can be studied over time irrespective of the changing nature of class membership. The nature of the work students produce can similarly be assessed over time with different individuals. Perhaps, then, one way of improving the feasibility and quality of action research in the UK EFL sector is to focus from the outset on topics and investigative strategies that do not assume a consistent sample will be available for several weeks. In university pre-sessional programmes this will be less of an issue, but it is a feature of most EFL schools in the UK.

Secondly, the problems created by sample attrition and instability are exacerbated when teachers adopt research designs which involve the quantitatve analysis of causal relationships. Lindsay Warwick's study, for example, was very carefully planned and conducted. However, the small number of students she worked with, and the inevitable lack of control she was able to exert over a range of variables, meant that her statistical results were inconclusive. Similarly, Flynn and Newby's attempts to study the relationship between regular self-assessment tasks and improvements in learner autonomy were limited by the fact that they were only able to retain three learners for each phase of the study. In both these cases the teachers collected qualitative data through interviews to supplement their quantitative measures, yet the latter seemed to carry more weight in shaping the overall findings in these studies. This does not mean, of course, that the process of doing the project did not enhance in valuable ways teachers' understandings of their teaching and their learners; it does mean, though, that answers to questions about whether a particular intervention has a particular result will always be inconclusive.

To summarise my thoughts, then, what I am saying is that rolling enrolment imposes significant limitations on sampling which make it difficult for teachers in the UK EFL sector to work with the same group of students over an extended period of time. This challenges one of the core principles of action research – the idea that teachers can improve understanding and educational practice by going through repeated cycles of intervention, evaluation and reflection over time. A desire to work with a consistent sample also typically means that this ends up being small, and this makes it difficult for action researchers who adopt quantitative pre- and post-measure designs to reach firm conclusions.

What this suggests then is that in the UK EFL sector action research will be more productive when it is designed in a manner that allows teachers to involve all of their students

(or at least all the students in a class), irrespective of the duration of their study period. It would also seem that in such contexts there are particular benefits to including a qualitative element (e.g. by talking to learners or describing their learning in more narrative ways) as well as more common quantitative measures.

Action research and teacher motivation

One indisputable fact about action research is that, irrespective of how conclusive results are, teachers always find the process extremely rewarding. Furthermore, when we ask teachers about how action research has affected them professionally, they talk less about specific results and more about the impact of the process on their motivation, enthusiasm, confidence, awareness of their learners, and criticality regarding their own assumptions and practices (Edwards and Burns 2016 and Goodall, Day, Lindsay, Muijs and Harris 2005 provide related insights into the impact of continuing professional development on teachers). These are powerful outcomes, so powerful in fact, that an argument could be made that this is where the true sustained value of action research lies; not in generating clear-cut results, but in providing the kinds of professional reinvigoration and attitudinal realignment that will stay with teachers long after the formal conclusion of any particular action research scheme. This is perhaps, too, how I hope this collection of papers will impact on readers; the projects do provide interesting insight into a range of key aspects of ELT, but what they provide above all is inspiration to make readers want to start examining teaching and learning in their own classrooms.

References

Borg, S (2015) Professional development through the Cambridge English/English UK Action Research Scheme, *Research Notes* 61, 3–5.

Burns, A (2010) *Doing Action Research in English Language Teaching: A Guide for Practitioners*, New York: Routledge.

Edwards, E and Burns, A (2016) Language teacher action research: Achieving sustainability, *ELT Journal* 70 (1), 6–15.

Goodall, J, Day, C, Lindsay, G, Muijs, D vand Harris, A (2005) *Evaluating the Impact of Continuing Professional Development (CPD)*, London: Department for Education and Skills.

Our 2015 action researchers with Cambridge English and English UK representatives at the English UK Teachers Conference in November 2015 (L–R): Fiona Barker (Cambridge English), Huan Japes (English UK), Richard Flynn, Andrew Taylor, April Pugh, Christian Newby, Simon Borg, Ceri Thomas, Fiona Wattam, Lindsay Warwick.

The effects of correction and redrafting on low-level students' ability to self-edit

FIONA WATTAM COLCHESTER ENGLISH STUDY CENTRE

Introduction

Much research has been written about how we should deal with student errors in writing. I have been teaching EFL for over 20 years, but it seemed to me that whatever method I tried – different types of correction codes, no correction codes, pre-writing remedial work, post-writing remedial work – my students seemed unable to see the errors themselves before they handed their work in. I wondered if this was because they just assumed I would correct their errors, or whether they really could not see their own errors, or maybe they were unable to understand the root cause of the error. So I decided to do some research into how I could train them to analyse their texts to find their own errors and whether this would improve their writing in the long term.

I noticed that the students I taught on Cambridge English exam preparation courses often recorded all of their marked writing tasks carefully so that they could refer back to them each time they were set a new writing task. Mistakes they made with register, layout, forms of address and task achievement didn't seem to reappear on the second attempt because they had learned from where they had gone wrong the first time. But my General English students didn't seem to see connections between the texts I corrected and their future progress. This is what I wanted to address in my action research project.

Feedback on students' writing

Much research has already been published on error correction for writing. However, as Ferris states in a paper summarising 'The "grammar correction" debate in L2 writing', although there have been 'several decades of research activity in this area, we are virtually at Square One, as the existing research base is incomplete and inconsistent' (2004:49). Ferris refers specifically to an earlier, well-known essay which argued that 'grammar correction has no place in writing courses and should be abandoned' (Truscott 1996:327). In this paper, Truscott argues that research has shown grammar correction to be ineffective and points out certain side effects 'such as its effect on students' attitudes or the way it absorbs time and energy in writing classes' (Truscott 1996:328). I was very conscious that the students would have to devote more time to their writing tasks than what they were used to and was interested to find out whether the students would see this as a good use of their time. Other researchers have argued against error correction feedback altogether, claiming that it does not take into account how language acquisition occurs.

Hyland and Hyland (2006:85) point out that because of the non-linear way in which languages are learned, we 'cannot expect that a target form will be acquired either immediately or permanently after it has been highlighted through feedback'. Ellis's (2009) typology of written corrective feedback types states that there are still 'no clear answers' to how teachers should correct second language students' writing and suggests that teachers may have to adjust the type of feedback they give depending on their learners (2009:106). In the face of so much diverse opinion, I wanted to find what worked for my students.

It is hard to imagine a student who would not want the teacher to correct their errors in some way. In terms of the types of correction, I did not set out to pit one against the other. In fact, I just wanted the simplest method possible of getting students to become aware of their errors. However, I soon realised that because a lot of research has been done comparing the effectiveness of different ways of giving students feedback, my choice of feedback could also have an impact on the students' progress. Chandler (2003) compares different types of teacher feedback, including Direct Correction (the teacher writes the correct form), Underlining and Description (the teacher underlines and uses a correction code to identify the type of error) and Underlining without Description. Chandler found that Direct Correction and simple Underlining were much more effective than describing the error with a code for 'reducing long-term error' (2003:268) but I decided not to use Direct Correction, as it was my firm belief that learners needed to engage with their errors as part of the learning process and I wanted them to develop the skill of working out their own errors.

But what Chandler felt was the 'crucial factor' in all of this was 'having the students do something with the error correction besides simply receiving it' (2003:293). This was also the crucial factor in my research. Hyland (1990:279) reiterates the argument that 'we must persuade students to act on the feedback we provide'. He suggests that teachers should not give a grade until the students have attempted to self-correct (1990:281) and Kozlova (2010:97) says that marks should only be given after redrafting, otherwise students will 'become more passive and abandon any further effort'.

Research design

Based on my context and the literature, the main questions my action research set out to investigate were:

1. To what extent does grammar correction through the use of correction codes and redrafting help students identify and self-edit their errors in the *short term*?
2. To what extent does grammar correction through the use of correction codes and redrafting help students identify and self-edit their errors in the *long term*?
3. Would the students view the redrafting and notebook system as beneficial enough to justify the extra workload for the teacher and student?

Context and participants

A large number of students who study at Colchester English Study Centre, a private language school where I work, do so in order to take the *IELTS* exam and go on to study on higher education courses. They often start at Common European Framework of Reference for Languages (CEFR, Council of Europe 2001) Level A1 and have to move up through the levels by passing a level test before they are ready to enter an *IELTS* preparation class. In these lower levels, the emphasis during lessons is often on developing spoken communication skills to help them survive in an English-speaking environment, so the written component of the level test is often a struggle to pass. Also, the school runs a continuous enrolment system, which means that the class profile changes regularly, as does the class teacher. Therefore it can be difficult for the teacher to get to know their students' writing abilities and to develop a method of correction and feedback that will address students' writing problems systematically. I wanted to give writing more importance so that I and other teachers would devote more time to writing in class rather than setting it for homework.

With this in mind, I selected a class of pre-intermediate students who agreed to keep all their written work in a notebook that I provided for them. They also agreed to rewrite every piece of corrected work, and to complete questionnaires at the start and end of the project. During the study, which lasted eight weeks, a number of students joined and left the class. Although everybody in the class was given the opportunity to redraft their writing after correction, and to complete questionnaires, I analysed the results of six core students who did exactly the same pieces of writing, so that I could compare their performance more consistently.

The six core students were Danilza (Angola), Dora (Taiwan), Bahar (Turkey), Hasan (Turkey), Mohammad (United Arab Emirates) and Jose (Brazil). Their ages ranged from 20 to 27, and they were full-time students at the school, studying for 25 hours a week. I taught them for 2 hours every day. One of the other students in the class found the rewriting part of the agreement a little challenging – I will refer to him as the 'control' student as he attended the same lessons and did the same writing tasks but did not redraft or keep them in a book.

I had already decided not to give students a grade and explained my reasons to the students – I wanted them to focus on the process of writing rather than the product. The school currently has no system for grading students' written work – teachers tend to write comments such as 'Good'/'Very good' as a means of encouragement rather than giving students a clear indication of their level. Instead of giving a grade, in my brief comments at the bottom of their writing I tried to be specific about what was positive ('Good use of vocabulary') and what needed work ('Try to use linking words – and, so, but, because').

Cycle 1 – Initial questionnaire

I decided that the simplest way to get information from the students was by using a Likert scale questionnaire with a total of eight questions to gauge their attitude towards how their writing was marked in the school and what they did with their writing once it had been returned (see Appendix 1). I gave the questionnaire to a total of 12 students who took part in the project at various starting points. I did not ask them to put their names on the questionnaire. The school already had a correction code system in place, which some of the students were familiar with, but there was no way of knowing whether students acted on the feedback or learned from it.

Initial findings

According to the results of the questionnaire, students were very positive about the use of the correction code system of marking that they had experienced at the school (Figure 1). All 12 students said that they looked at the errors and correct them (Figure 4), with most students referring back to previous writing before handing in a new writing task (Figure 7). I found this quite surprising, given the lack of emphasis on writing skills for lower levels at the school. It is possible that social desirability was a factor here – students were possibly saying what they thought I wanted to hear, as there is a tendency for respondents to give answers that reflect acceptable behaviour rather than what they actually do. Over half of the students said that the use of the correction codes had helped them to improve their writing (Figure 8) and over half did not want the teacher to write the correct form next to their error (Figure 3). However, only three students said that they keep their writing tasks in a notebook (Figure 6), so this was definitely an area that needed more research. Two thirds of the students (Figure 2) said that they were unsure about how to correct their mistakes, and this is something which I tried to address at the start of the project using a worksheet which gave students practice in using the code that I was planning to adopt (see Appendix 2). Only one person thought that rewriting the same text that had already been corrected was a waste of time (Figure 5). Perhaps this was the one student who did not hand in any redrafted versions.

Figure 1: Question 1

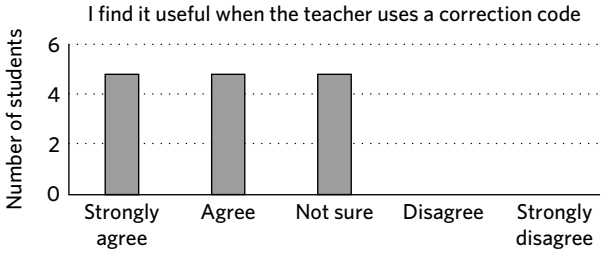


Figure 2: Question 2

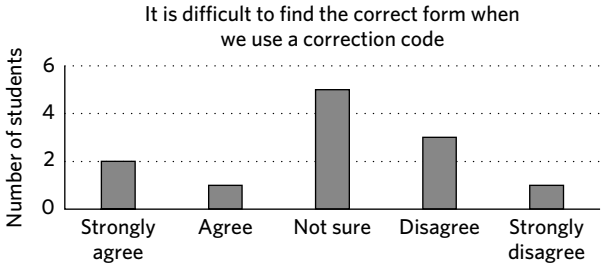


Figure 3: Question 3

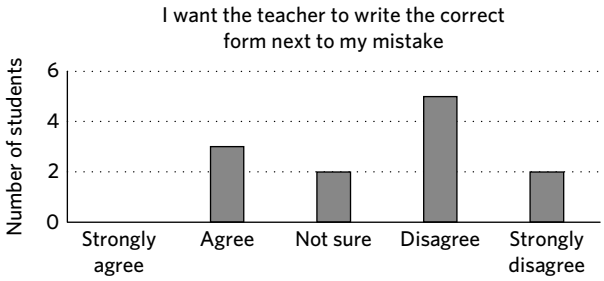


Figure 4: Question 4

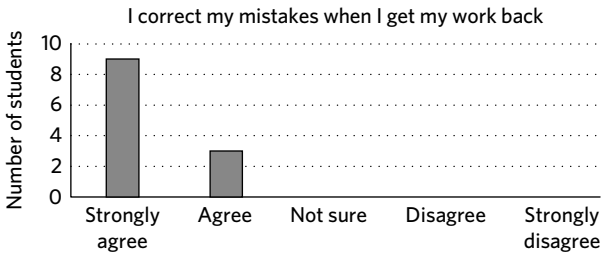


Figure 5: Question 5

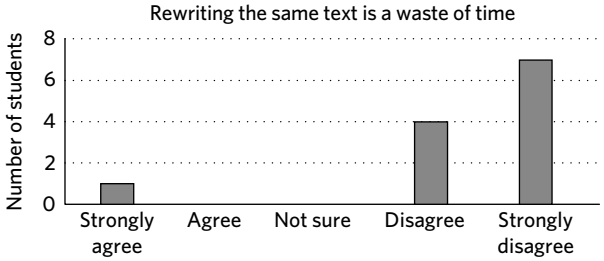


Figure 6: Question 6

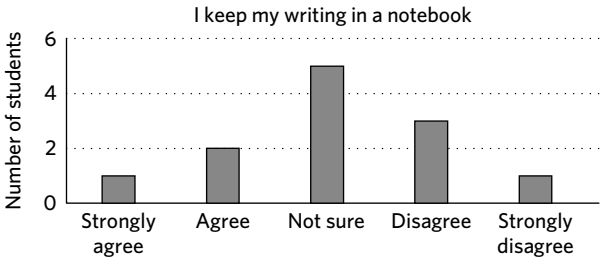


Figure 7: Question 7

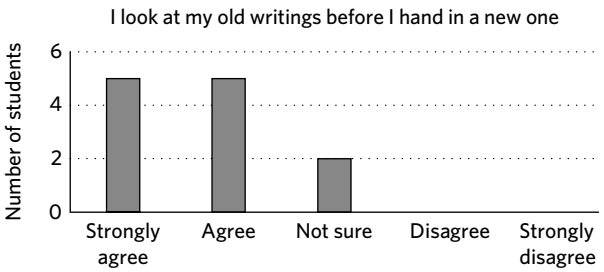
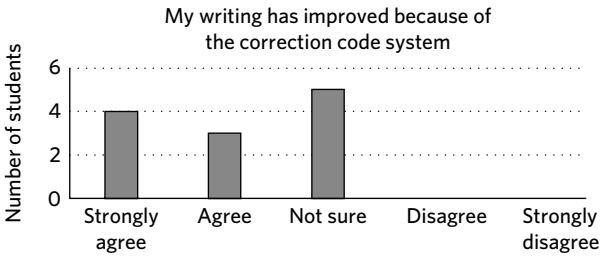


Figure 8: Question 8



Choice of correction code

I decided to keep my correction code as simple as possible. I just wanted to see what would happen if the learners corrected their errors consistently and regularly, and to limit my corrections to only language that they had studied or were studying (e.g. articles, past tense forms, singular/plurals, pronouns and spelling patterns with general applicability). I wanted to focus on errors that I considered to be 'treatable' (Ferris 1999:6) or that might 'stigmatize the learner' (Makino 1993:337), for example handwriting. Although handwriting was not something I addressed directly in my feedback, it did become one of the factors that I took into account when measuring individual progress in specific areas.

I decided to mark the error in a way that would make the meaning as clear as possible. For example for the missing third person 's', I marked this as 'Something missing' rather than 'Form', so that the student would get used to putting something at the end of a third person verb. I also limited my 'Wrong word' correction to errors that are appropriate for their level e.g. 'homework'/'housework', 'a cook'/'a cooker'. Finally, I made a brief comment at the bottom of the text that focused on a key area to work on.

After the questionnaires, students did four writing tasks of different genres – one per week (see Table 1). I gave

the writing task on a Thursday and students completed it during class time. I marked it the same day, spent some time in class on Friday helping individuals if they had problems with deciphering the code, and then gave them the weekend to produce a corrected version next to the original version in their notebook.

The choice of genres was linked to the coursebook we were using at the time, so that the whole process was integrated into the weekly topic and grammar focus. The genres were also typical of the type of writing they might need in real life (letters and emails) and for their level tests (descriptions) and future exams (opinion essays). Support was given by using the context presented in the coursebook to help students with vocabulary and grammatical forms together with a brainstorming of ideas and an analysis of a model. Students were then given 20 minutes to produce a text of about 150–200 words in class without using dictionaries or asking for help. I decided to use these test conditions in order to ensure that students were producing their written work under comparable conditions, which would allow me to rule out different conditions as an explanation for any differences in students' performance. I also wanted to promote writing as something to be done in class rather than given for homework.

Table 1: Writing tasks in Cycle 1

Week (Task number)	Task
Week 1 (Task 1)	'My single friend' – write a description of your friend describing his/her appearance, job, hobbies and interests.
Week 2 (Task 2)	Letter to the council – write a letter to the local council complaining about a new transport scheme that you have read about in the newspaper.
Week 3 (Task 3)	Email to a classmate – write an email to a classmate to arrange a day trip together on Saturday.
Week 4 (Task 4)	Essay – Do you think footballers are paid too much? Write the reasons why you think they should or shouldn't get a high salary and say which jobs you think should be paid more.

'The Blue Book'

Students were given slim, A4-sized notebooks with a plain blue cover. Students were encouraged to take ownership of the notebook by decorating the cover in any way they wished. The correction code was glued inside the cover for easy reference. I discussed with the students the following reasons why the notebooks were important for the project:

1. Having all their writing in one place would make it easier for them to look back at previous writings and see mistakes they had made before.
2. The notebook would encourage students to be neat and organised in their work.
3. The notebook would help the teacher and student monitor progress over time.
4. When the students changed teacher, group or level the new teacher would quickly be able to identify areas to work on.
5. The notebook would act as a 'portfolio' of work, which would provide evidence of ability when the student is ready for the next level.
6. Students would always have models to refer to when producing texts of a similar genre.
7. Students could take the notebook with them as a memento of their stay.

One of my initial concerns about the project was that students would lose the books and that all their work would be lost. I considered keeping all the books at school but I thought this would remove the idea of taking responsibility for their own work. Fortunately none of the books was lost despite one of them being left on a bus and subsequently retrieved.

Cycle 2

After four weeks of focusing on errors being pointed out clearly to the students by the correction code, I wanted to see if this had started to have an effect on how they were seeing their own writing and what difference the feedback had made to how easily they could identify and correct their errors.

I decided that in the second cycle I would indicate the errors by underlining. This time my aim was to observe how capable they were of correcting errors after they had had four weeks of errors being clearly pointed out to them.

To better achieve this, I took a different approach, which was to sit next to the learner during class time after I had underlined their errors, and ask them to correct the mistakes orally and to talk me through their thought processes. This approach allowed me to ask questions about how much the student knew about the mistake they had made (e.g. choice of 'a'/'the' article) and the reasons for their choice. I followed a similar genre pattern to Cycle 1, again working around the coursebook topics (see Table 2). At the end of this cycle I set a final writing task (Week 8) and analysed this in terms of number of errors as a percentage of word count.

Table 2: Writing tasks in Cycle 2

Week (Task number)	Task
Week 5 (Task 5)	Description of a restaurant – write a review of a restaurant that you go to regularly – describe the food, décor, service and atmosphere and say why you like it.
Week 6 (Task 6)	Letter to a host family – write a letter to your new host family telling them about your travel arrangements.
Week 7 (Task 7)	Email to a friend about a story – tell your friend about something funny or interesting that happened to you in Colchester.
Week 8 (Final task)	Essay: Which is better? Swimming in the sea or swimming in a public swimming pool? Write about the benefits and drawbacks of both, and say which one you prefer.

Findings

I will discuss overall findings and then look at individual cases.

Cycle 1 – changes in the number of errors made in first drafts using a correction code

In this section I will present the findings which seek to answer the first research question regarding whether using a correction code helps students identify and self-edit their errors in the *short term*.

I counted the number of errors in Versions 1 and 2 of each task and then calculated the percentage of errors in relation to the number of words. I counted each time I marked the paper as one error, so every time I drew an arrow for something missing (articles, third person 's') or wrote 'Sp' for spelling or 'T' for tense counted as one error. As we will see later, the percentage of mistakes in Version 2 is lower. This gave me evidence that on each occasion, students had looked at their errors and corrected them in the redraft. However, the errors were not eliminated and there are two main reasons for this. Firstly, students often over-compensated by trying to write something new and different from their first attempt, which resulted in further error. Secondly, students did not always correct their mistake accurately, which suggests that they needed more time to process the reason for the error and also to break out of previous habits. My 'control' student, who never produced a redrafted version of his first attempt, consistently wrote 'Firstable' instead of 'First of all' despite my pointing it out with the correction code in three different writing tasks and discussing the error with him.

Over the four weeks, a very slight decrease in the percentage of errors in their first draft can be seen in most cases. This belies the fact that as I marked the texts, I saw a noticeable decrease in specific problem areas for individual students (see the section on 'Individual case studies'). Also, there was considerable improvement in areas that I was not able to measure by counting errors, such as handwriting, which I will discuss later. So in answer to the first research question, there was some evidence that, in the short term, students reduced the number of times they made certain types of mistakes in *subsequent* texts (as opposed to redrafted ones) after being made aware of them.

Cycle 2 – changes in the number of errors in the first draft with less support

The aim of Cycle 2 was to investigate the second research question – to see if the correction and redrafting system would improve their ability to self-edit and improve their writing in the *long term*. For Tasks 5, 6 and 7, I asked students to produce a redrafted version of their first attempt, but I did not collect data for this cycle as my aim here was to spend time getting to know how they handled the errors after four weeks of being trained to self-edit with a correction code. In order to give less direct indication of error, I used underlining to highlight the location of the error, and then discussed the errors with the learners individually to see how easily they could self-correct.

During this period I noted some of their reactions to the errors and how they accounted for them:

Jose:	'Yes, I know, I always make this mistake. In my language, is different' (referring to the omission of 'it' in phrases such as 'It's difficult').
Teacher:	'If you know it's different in your language, will this help you remember it next time?'
Jose:	'Maybe in writing, yes, but in speaking it sounds the same so I forget in writing.'
Danilza:	'I don't know where to put full stop. If I use commas the sentences are longer so it's better.'
Mohammad:	'Nobody told me I have problem with handwriting. Now it looks better.'

What struck me most in this cycle was how readily the students could correct their errors having got used to correcting specific problems highlighted by the correction code. They easily identified missing articles and the omission of the third person 's', could quickly correct punctuation, usually saw tense and form problems but had problems with correcting spelling and wrong words, for example using 'make' instead of 'do'. I could see from their quick corrections that their error had not come from an underlying misunderstanding of the grammar. The usual response was 'Oh yes I know this!' but when I asked them why they still made the mistake it was difficult to get a clear answer. This perhaps demonstrated that they were still processing new language and that the learners just needed more time for the intervention to have a lasting effect.

As mentioned previously, the aim of the final text was to answer the question of whether the intervention would help improve their ability to self-edit in the long term. Looking at the graphs (see Figures 9–14), the percentage of errors in the final task seems to revert to where it was in the first task in most cases.

Figure 9: Danilza's errors in task 1-4 and final task

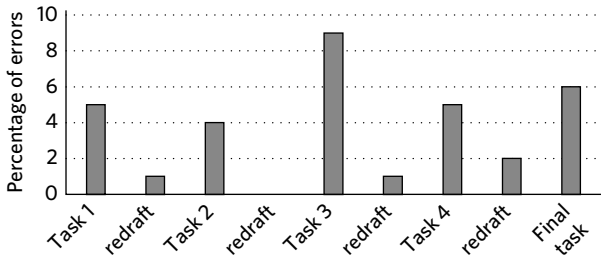


Figure 10: Hasan's errors in task 1-4 and final task

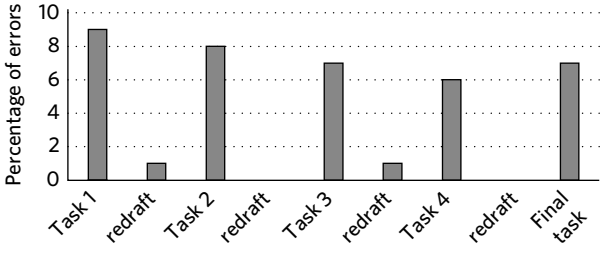


Figure 11: Bahar's errors in task 1-4 and final task

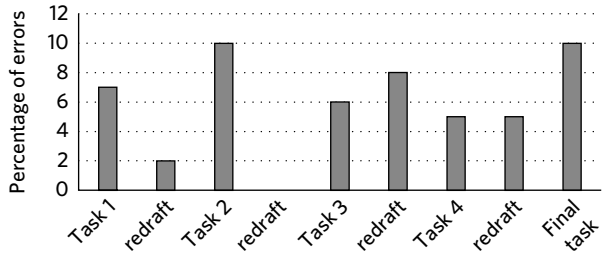


Figure 12: Jose's errors in task 1-4 and final task

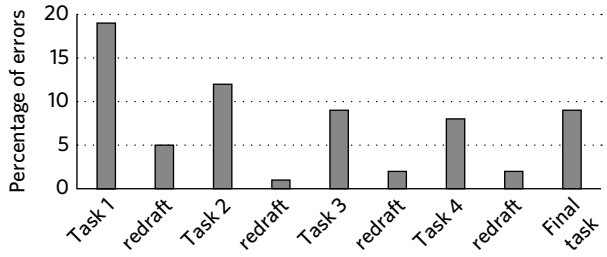


Figure 13: Dora's errors in task 1-4 and final task

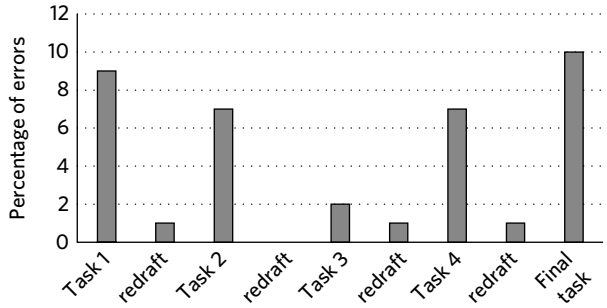
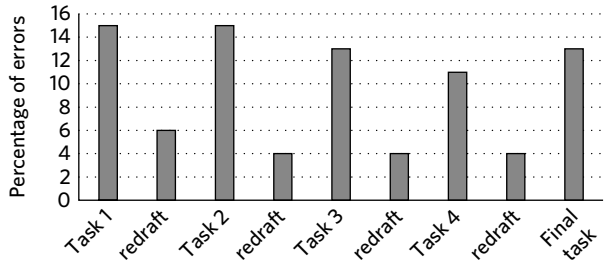
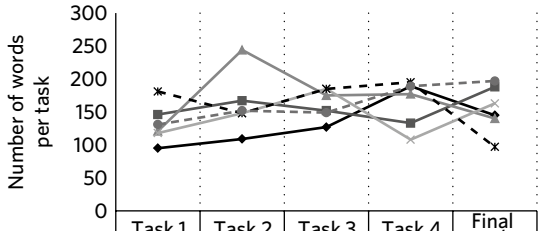


Figure 14: Mohammad's errors in task 1-4 and final task



Some studies suggest that a focus on error correction leads to a reduction in fluency (number of words) and complexity (range of vocabulary and structures). Interestingly over the period of the study, the number of words gradually increased and though I have not analysed the complexity of the students' final writing in relation to the first, one explanation for the increase in errors was that they were taking more risks, feeling more confident, and enjoying their writing. The changes in the number of words produced can be seen in the chart below (Figure 15):

Figure 15: Increase in number of words per task



	Task 1	Task 2	Task 3	Task 4	Final task
Mohammad	95	109	127	189	145
Dora	146	167	152	133	188
Jose	121	244	175	177	140
Bahar	118	148	185	108	163
Hasan	181	148	185	195	97
Danilza	131	152	149	189	197

Findings: Individual case studies – Cycle 1 (Tasks 1–4) compared with the final task

As noted previously, although the overall number of errors increased on the final task, there were a number of noticeable improvements in specific areas of students' work. Ferris (1999:4) defines effective error correction as that which is 'selective, prioritized and clear' and says that there is evidence that this type of correction 'can and does help at least some student writers'. The results below showed that prioritising specific aspects of a student's writing and giving clear feedback could and did help them.

Mohammad had been studying at the school for nine months and had repeated pre-intermediate level several times. His main failing was writing and he had lost motivation to improve. However, he joined the project enthusiastically, seeing it as a final chance to see progress before he returned to his country. What was interesting about Mohammad was that though his redrafted version improved and his writing was more controlled and legible, he always reverted back to his previous way of writing when he started a new task. By the final task though, he was writing consistently on the line in a neat and organised style (see Appendix 3). He also appeared to start making progress with his weakest area, spelling (Figure 16).

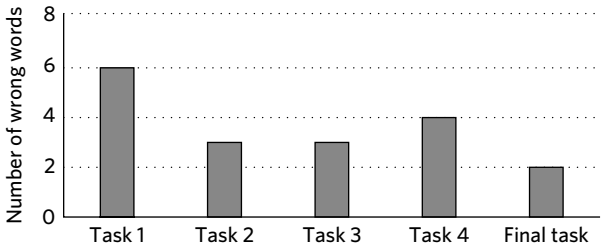
Figure 16: Mohammad's number of spelling errors

Jose struggled with the category of 'Something missing' (articles, plurals, third person 's' and prepositions) but in the final task he made far fewer mistakes than in the first one (Figure 17).

Figure 17: Jose's number of items left out

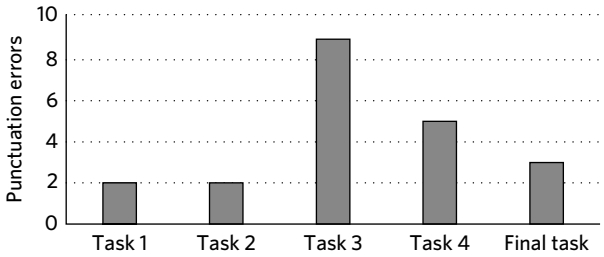
Hasan is an ambitious and highly motivated student who wanted to go on to study at university – for him, vocabulary was key to his desire to understand and express himself with sophisticated language. The high incidence of error at the start made him more aware of the importance of being selective about new vocabulary and using it accurately and appropriately (Figure 18).

Figure 18: Hasan's number of wrong words



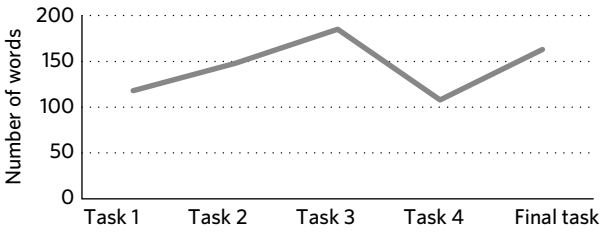
Danilza wrote confidently and fluently, but ended every sentence with a comma. It took her some time to adjust her way of writing so that she ended one sentence and started another. My intervention seemed to cause confusion and meant that she over-compensated and started putting full stops in all the wrong places. However, by the final test she seemed to have reached a better understanding of commas and full stops (Figure 19).

Figure 19: Danilza's number of punctuation errors



Bahar had the most inconsistent results in every area, with little evidence to show that any individual aspect of her writing improved. However, Bahar had always complained that she didn't know what to write about and this meant that her texts were very short and simplistic even though she always had plenty to say in class, a wide vocabulary and did well in grammar tests. In my written comments to Bahar I focused on areas where she could have said more – given more description, used more adjectives, and more detail such as reasons or explanations. This had a noticeable effect on how much Bahar produced in her writing, and as she was not used to writing longer texts, she had different types of error on different occasions. I would say that this was beneficial to her, as the more she produced, the more opportunity she had to deal with problem areas and this was part of the learning process for her (Figure 20).

Figure 20: Bahar's increases in number of words



Final questionnaire

Although there is not much evidence to suggest that the intervention had a huge impact on the students' ability to self-edit, on a subjective level, I felt that students were making good progress and were motivated to keep going. I never needed to chase anybody for their redrafts (apart from the 'control' student) and no notebooks were 'lost' in the process. The attitude of the students was positive throughout, and they even asked me to give them an extra piece of writing per week. Students who joined the class for short periods volunteered to re-write their first version, and students who moved up to a different class asked if they could continue with the notebook system.

But there were a number of variables in the project and I was keen to find out what students thought had been the most beneficial aspect. It is difficult to determine what exactly helped the students – whether it was the increased amount of time devoted to writing in class, the redrafting, the notebook, or the consistent use of a correction code. I was surprised by the results of the final questionnaire (see Appendix 4), as it brought to my attention something that I had not taken into account and brings us to the final research question (would the students view the redrafting and notebook system as beneficial enough to justify the extra workload for the teacher and student?).

Most students agreed that the project had helped them notice and understand mistakes in their writing (Figure 21 and 22), though not everybody thought it had helped them to stop making mistakes (Figure 23). Only one person said that it had not improved his writing (Figure 25) and I was quite surprised that some students had not found the correction code easy to understand (Figure 24). All of the students strongly agreed that they would like to continue with the project (Figure 26).

Figure 21: Question 1

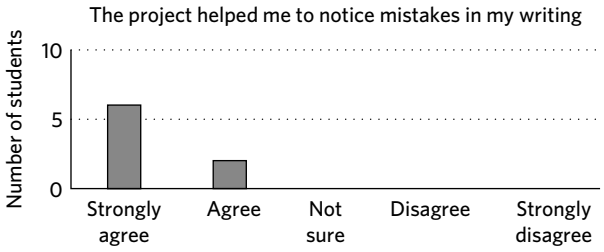


Figure 22: Question 2

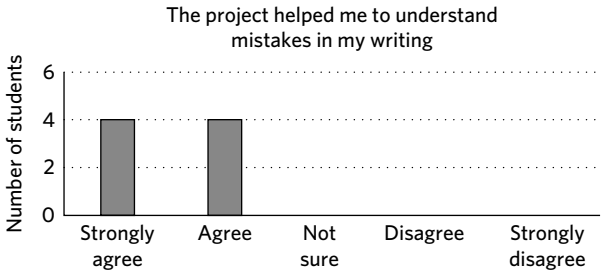


Figure 23: Question 3

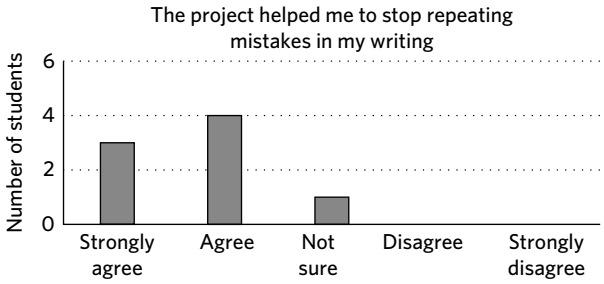


Figure 24: Question 4

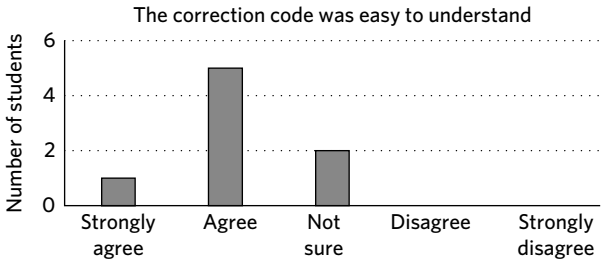


Figure 25: Question 5

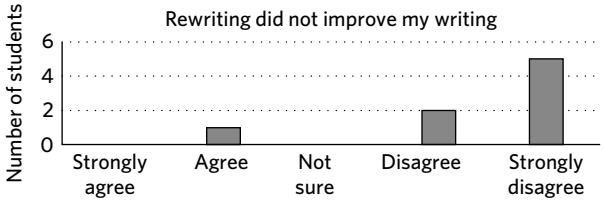
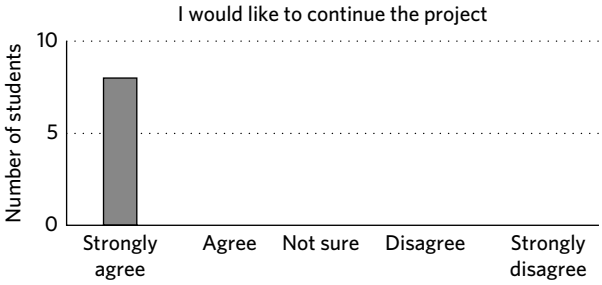


Figure 26: Question 6



Finally, the questionnaire asked students to underline the one thing they found most useful. The results are in order in Table 3.

Table 3: What students found most useful

Most useful aspect of project was . . .	Number of students
. . . the teacher's comments at the end of the writing	5
. . . writing regularly	2
. . . using the correction code	1

I was surprised to find out that actually, the two aspects of the project that I considered most important (rewriting, keeping all writing in a book) were the least important to the students, and the thing that stood out as being crucial to students was the comment I wrote at the bottom of the writing.

Three of the six students explained their choices as follows:

- The teacher's comments at the end of the writing:
  'When you speak with a teacher about the mistakes is easier understand different things. All of this tips are very important, but for me, to understand my mistakes was the comments at the end of the writing.' (Jose)
- Writing regularly:
  'Because I'm going to enter the *IELTS* exam so I needed to write an essay regularly.' (Hasan)
  'In my opinion, practice more is the best way to improve English (if I stop to practice, I would forget the grammar gradually. Also, writing similar topic could be helpful for me to describe things through different way.' (Dora)
- Using the correction code:
  'Because it make me understand my mistakes and help me to improve my writing.' (Danilza)

Reflections

I had set out to find a way of improving students' writing that would give me some confidence in knowing that I was doing the right thing. I also wanted to see the positive results of my Cambridge English exam classes replicated at lower levels. I felt frustrated that lower levels made the same mistakes repeatedly, and wanted to see if I could change this.

I have no doubt that the notebook system gave added value to the students' course. The fact that they were so keen to continue with the project showed that they had enjoyed it and saw value in it. What surprised me was how little time it took to mark the work once the students had become familiar with the correction code system. Underlining errors took very little time, and because I could easily refer back to previous writings, a short comment on progress (or lack of) in one particular area was no hardship, and much more useful to the student than a 'Good' comment.

With regard to the action research project, it has helped me to look at what I do in class with a more critical eye. I will continue to question students on how they feel about what I do in class and to give them the opportunity to sample different techniques. I will always try to find out which techniques were more effective and which ones the students preferred. Action research is something that teachers can apply to any aspect of their teaching and I guarantee that there will be surprising, interesting and informative learning points for all those involved. The project has also reminded me that there is a lot of research out there and that it should inform what we do. Finally, it has made me realise how important it is to ask students what they think – so often we're just afraid of getting an answer that we don't want to hear.

References

Chandler, J (2003) The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing, *Journal of Second Language Writing* 12, 267–296.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.

Ellis, R (2009) A typology of written corrective feedback types, *ELT Journal* 63 (2), 97–107.

Ferris, D R (1999) The case for grammar correction in L2 writing classes: A response to Truscott (1996), *Journal of Second Language Writing* 8 (1), 1–11.

Ferris, D R (2004) The 'grammar correction' debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime. . .?), *Journal of Second Language Writing* 13, 49–62.

Hyland, K (1990) Providing productive feedback, *ELT Journal* 44 (4), 279–285.

Hyland, K and Hyland, F (2006) Feedback on second language students' writing, *Language Teaching* 39, 83–101.

Kozlova, I (2010) Ellis's corrective feedback in a problem-solving context, *ELT Journal* 64 (1), 95–97.

Makino, T (1993) Learner self-correction in EFL written compositions, *ELT Journal* 47 (4), 337–341.

Truscott, J (1996) The case against grammar correction in L2 writing classes, *Language Learning* 46 (2), 327–369.

Appendix 1: Initial questionnaire (my writing)

When the teacher marks my writing . . .	Strongly agree	Agree	Not sure	Disagree	Strongly disagree
1. I find it useful when the teacher uses a correction code such as 'Sp' for 'spelling'.					
2. When the teacher uses a correction code with my writing, it's difficult to know what the correct form is.					
3. I want the teacher to write the correct form (**not** the correction code) next to my mistakes.					
4. When the teacher gives me my work back, I look at the mistakes and then correct them.					
5. Re-writing the same text, after I have corrected the mistakes, is a waste of time.					
6. I keep my corrected writing in a separate notebook.					
7. I look at my old writings before I hand in a new one, so that I don't make the same mistake twice.					
8. My writing has improved because of the correction code system.					

Appendix 2: Writing correction code

Look at the teacher's corrections in the writing below. What do they mean? Discuss with your partner.

P F T
i like study at CESC. I studied here for six monthʌ. Colchester is
ʌ very intresting place to inhabit.
Sp WW

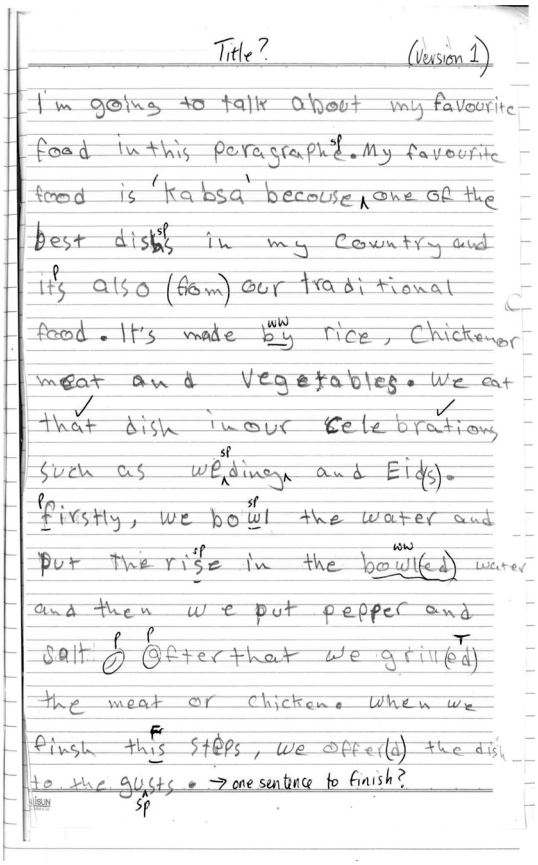
My class is very international. The other students (are) come from

many countries different. We get over very much.?
WO P

Table 4: Table of typical codes and errors

Code	Meaning	Some examples
P	Punctuation	**C**apitals, small letters, full stops(.) commas (,)
F	Form	Gerund/infinitive – *I like studying*
T	Tense	Past tense, Present tense
ʌ	Something missing	Articles (a, the), prepositions, plural 's', third person 's'
Sp	Spelling	Double or single 't' – *writing/written,* y – i: *I stud**y**, he studies*
WW	Wrong word	Wrong translation, false friend, too formal
WO	Word order	Adjectives before nouns – *many different countries*
?	I don't understand	
( )	You don't need this	

Mohammad - Excerpt from first piece of writing, showing handwriting



Appendix 4: Final questionnaire

Dear student

Thank you for taking part in my writing project. I would like to find out what you think about the project. Please could you complete the table below and write any extra comments in the space at the bottom.

Fiona

	Strongly agree	Agree	Not sure	Disagree	Strongly disagree
1. The project helped me to notice mistakes in my writing.					
2. The project helped me to understand my mistakes.					
3. The project helped me to stop making the same mistakes.					
4. The correction code was easy to understand.					
5. Re-writing the same text, after I corrected the mistakes, did not improve my writing.					
6. I would like to continue the writing project.					

What was the thing which you found **the most useful** about the writing project?

(Please underline **one** only)

- Using the correction code
- Rewriting each task
- Writing regularly
- Keeping all my writing in a book
- The teacher's comments at the end of the writing
- None of these

Please explain the reason for your choice (you can write on the other side if you want to)

The use of assessment criteria in classroom speaking tasks

LINDSAY WARWICK BELL CAMBRIDGE

Introduction

Over the last two years at my school, Bell Cambridge, we have been exploring the use of formative assessment tools to help our students' progress. While researching this, I read the results of Black and Wiliam's seminal meta-study *Inside the Black Box* (2001), which showed that formative assessment can have a significant, positive effect on learning and I very much became interested in knowing more. Further research revealed that Wiliam (2011:46) believes there are five key areas of formative assessment: discussion and tasks, effective feedback, the sharing of learning intentions and success criteria, learning from other learners, and students taking control over their learning.

Four of these areas were clear to me but the term *success criteria* was not and I wanted to know more. I was interested to learn that it described a set of criteria given to students before a task that tells them how they can be successful in that task and ultimately achieve the intended learning objective. The criteria are regularly used in state schools and are also used for self- and peer-assessment after a task, but in my opinion they are not generally used in EFL lessons except perhaps in exam classes and to a loose degree General English writing lessons. Having rarely provided students with this kind of criteria myself, I was really interested to know if they could help my international adult students at a private language school in the UK. In particular, I wanted to examine the use of success criteria in relation to self-assessment and the development of speaking skills.

Background

Self-assessment is generally considered to be a positive influence on student learning and encourages students to develop metacognition. McMillan and Hearn (2008:48), having reviewed both theoretical and research papers, said that 'when students set goals that aid their improved understanding, and then identify criteria, self-evaluate their progress toward learning, reflect on their learning, and generate strategies for more learning, they will show improved performance with meaningful motivation'. They also believe that self-assessment gives students the opportunity to internalise the criteria against which they can measure their own success.

Looking at ELT-specific research, there have been many small-scale studies into the relationship between criteria and the ability to self-assess and/or improve speaking performance. Much of this research has found that the relationship is a positive one. For example, Chen (2008) conducted a study with 28 university students taking oral training classes. Students self-assessed their speaking against criteria and completed an evaluation form and questionnaire. The criteria were created by teachers in collaboration with students and students were given training in how to use them for assessment. The study also incorporated teacher assessment and peer assessment. The researcher concluded that self-assessment is 'both a viable alternative to teacher assessment of oral performance and a useful learning task' (2008:255); as well as the fact that students under-marked themselves at first but their marks became more in line with those of their teacher and peers with practice and feedback.

A study by Babaii, Taghaddomi and Pashmforoosh (2015) found that when 28 university students had agreed upon assessment criteria before they completed the task and had some practice, the difference between their assessment and the teachers' assessment was narrower. They also found that students perceived the use of self-assessment to be generally positive. Huang and Gui (2015:129) carried out a study among 61 college students in China, which showed that when those students received assessment criteria before they completed a speaking task, their speaking improved in terms of discourse length, organisation and linguistic flexibility, although the students did not improve in accuracy of pronunciation and grammar or range of tense usage. Students also believed that the criteria were a useful tool for self-assessment.

What exactly constitutes assessment or success criteria is a more complex area. In state schools in the UK, success criteria tend to comprise a list of instructions whereas in EFL an assessment rubric or scoring rubric is used, for example in *IELTS*. Litz (2007) defines an assessment rubric or scoring rubric as consisting of 'a fixed measurement scale and a set of criteria that are used to discriminate among different degrees of quality or levels of proficiency'. A rubric is more in depth than a list of instructions with very clear guidance as to what is required by students to meet each criterion in each band.

Research questions

The purpose of my study was to find out if the use of assessment criteria could help students develop both their speaking skills and the ability to self-assess their speaking skills. As a result, my two research questions were:

1. Does the presentation of assessment criteria before a speaking activity help learners to self-assess their speaking skills more accurately?
2. Does self-assessment result in an improvement in speaking skills?

I wanted to focus on speaking activities that students may do at the end of a lesson or end of a series of lessons that give them the opportunity to use language they have studied during the lesson(s). These may be considered as *production* activities.

Participants

I carried out this action research project with a class of pre-intermediate (Level A2 on the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001)) level students over a 7-week period. Nine students took part in the study, of which seven were Arabic speakers, one was a Chinese speaker and one was a Thai speaker. Eight of the students were male. They were aged between 18 and 43 and were studying English full time in the UK to prepare for an English-speaking university course or for their work. All students were given a letter outlining the details of the research and agreed to participate in the study. Four students completed all six speaking tasks; the other five completed five tasks. This was for reasons of illness, holiday or rolling enrolment as two students joined the class in week two of the study.

The study

The study comprised two phases. Each phase involved three speaking activities, one towards the end of each week over a 6-week period. Students were given the task and allowed 5 minutes to prepare. They then worked in pairs or small groups and completed the task which was recorded on a mobile device belonging to the teacher. Students were paired or grouped differently each time and did not necessarily speak in the same order. The speaking tasks can be found in Table 1.

Table 1: Study speaking tasks

	Phase 1	Phase 2
Task 1	Describe someone in your family.	Say why an invention is more important than other inventions on a list.
Task 2	Talk about if your life is easier or more difficult than when you were young.	Describe a wedding in your country.
Task 3	Tell a story from picture prompts.	Talk about a wish you have.

In the first phase, *after* students had completed the speaking task, they completed an assessment form (see Appendix). The first time, we carefully discussed the form and what each criterion meant together. Terminology was taught. After this, students completed the form on their own. The form was collected, I listened to the recording and completed a copy of the same assessment form with my own assessment of the students' work, without looking at the students' self-assessment. I gave written comments at the end outlining key strengths, an area to work on and advice on how to do this. Students then received a copy of both assessment forms to read and compare in class.

In the second phase, the procedure was the same as above except that the students were given the assessment criteria *before* they did the speaking task. Some of the criteria were given but the task completion and language use/accuracy sections were blank so we could discuss together as a class what kind of content and language they should include to complete the task successfully. The criteria were completed on the board and on a paper copy. The latter was copied so each student had a copy before they did the task.

Originally I intended to include four or five instructions for each task (e.g. 'talk about appearance and character', 'use comparative adjectives') and ask students to say if they had done each thing well, quite well or not well. However, having considered examples of assessment rubrics used in EFL research, such as those used by Huang and Gui (2015), I decided to use a more in-depth rubric instead to ensure consistency across all six tasks. The rubric (see the Appendix) focused on five key areas of speaking each time: *task completion*, *fluency*, *language use*, *language accuracy* and *pronunciation*. 'Language use' referred to range of language used and was aimed at encouraging students to use the language point studied in class that week. These criteria/categories were chosen as they are assessment criteria common to both the Cambridge English suite of exams and the *IELTS* exams. As most of my students were intending to take one of these exams at a later date, I felt that it would be good for them to become used to these criteria. Students were required to rate themselves as having achieved each criterion: well, quite well or not well. They were given further information about what each grade meant e.g. to achieve 'I did the task well' in 'task completion' for Task 2, students needed to 'talk about past and present obligations'; or to achieve 'I did it well' in 'language use', students needed to use '(don't) have to/(didn't) have to'.

Students were also asked to give themselves an overall score out of 10 for their performance. The purpose was to allow students to note overall progress and allow further comparison of the teacher's assessment and the student's assessment. A scale of 1–10 was used in order to allow small levels of improvement to be shown, as the study was conducted over just seven weeks. With hindsight, this grading was unnecessary as I was able to calculate an overall score from the five ratings for each criterion.

The self-assessment questionnaires were analysed, with the quantitative data being added to a spreadsheet. The teacher's assessment was similarly added so that a comparison could be made with regard to 1) the assessment criteria for each of the six tasks, and 2) the overall rating out of 10 for each of the six tasks.

As well as drawing on the quantitative data described above, I conducted and recorded an interview with each participant in Week 7 of the study, after the speaking tasks had all been completed. The purpose was to explore their experiences and opinions of the use of assessment criteria and self-assessment in their speaking development. Each interview was between 5 and 10 minutes long. I chose to conduct semi-structured interviews as these provide both structure and the freedom to 'follow up interesting developments and to let the interviewee elaborate on certain issues' (Dörnyei 2007:136). After the interviews were completed, a transcript was made which was then analysed for common experiences or beliefs.

Key findings

Assessment criteria comparison

I will begin by focusing on the ratings that I and students gave in relation to the five criteria: task completion, fluency, language use, language accuracy, pronunciation, i.e. I did it well, I did it quite well, I didn't do it well.

First, I looked at the ratings that students gave themselves for each criterion in each task and compared these with the teacher's rating for the same criterion. I wanted to know how often the students gave themselves the same ratings as the teacher, how often they gave themselves a higher rating and how often they gave themselves a lower rating. The results are set out in Table 2.

Altogether there were 114 individual ratings in Phase 1 and 120 in Phase 2. In 41% of cases in Phase 1, students and the teacher allocated the same rating. In Phase 2, this increased by 11% to 52%. In 48% of cases in Phase 1, students allocated themselves a lower rating than the teacher. This dropped to 39% in Phase 2.

Overall, these results show that students' ratings became closer to my ratings in Phase 2 when they were given the assessment criteria before they spoke. This means that students' accuracy in self-assessment did improve in Phase 2. However, it is important to remember that by Phase 2, students had self-reflected and completed the form three times. They had also received feedback from the teacher. This may have been the cause or at least have contributed to this outcome.

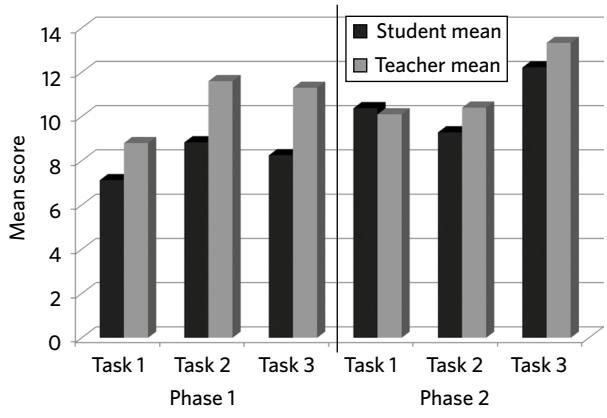
Table 2: Ratings comparison of students and teacher

	Total number of ratings	No difference between student and teacher rating	Lower rating by student	Higher rating by student
Phase 1 (criteria after task)	114	47 (41%)	55 (48%)	12 (11%)
Phase 2 (criteria before task)	120	62 (52%)	47 (39%)	11 (9%)

Second I looked at the mean ratings for each task so that I could compare them. I allocated each rating a score: 1 for 'I didn't do it well', 2 for 'I did it quite well' and 3 for 'I did it well'. I then calculated the average score per student per task. The results can be seen in Figure 1. In Phase 2, when students had the criteria before they did the task, their mean scores for all three tasks were higher than for the three tasks in Phase 1 which meant that they assessed their speaking performance as better in Phase 2. Similarly, the differences between the students' and teachers' scores were closer in Phase 2 than in Phase 1, reinforcing the results in Table 2.

However, Figure 1 also tells us that, while students rated their performance in the task as better in Phase 2, my mean scores rated their performance as lower in two tasks. This means that, according to the teacher, their performance improved in only one task (Task 3) when they had the success criteria before they did the task. This raises the question of whether students became more critical of their speaking, the more they self-assessed. It is possible that they did, although the fact that they rated themselves highly on the final task suggests this might not be the cause. Students did comment in class that Tasks 1 and 2 in Phase 2 were more challenging than the other tasks (See Table 1). Students felt that the level of ideas required of them was higher in Task 1 (technological inventions) and that Task 2 focused on a topic most knew little about (weddings). This may have impacted on performance.

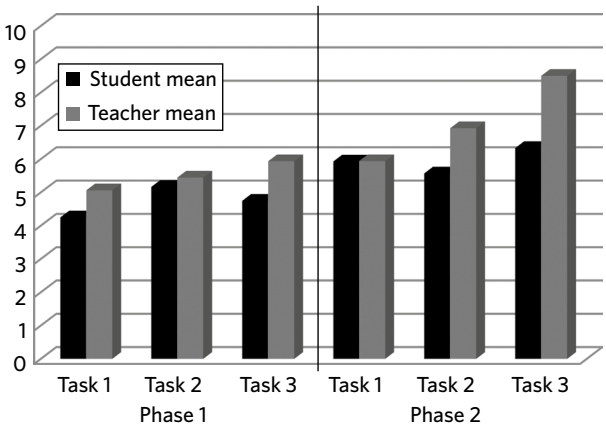
Figure 1: Mean scores per task for both student and teacher



Overall performance score

I will now focus on the rating out of 10 given by students and the teacher for the students' overall performance. The results can be found in Figure 2. In Phase 2, the students' mean overall performance scores were higher than those in Phase 1 when students had the assessment criteria only after they spoke. The teacher's mean overall performance scores were either the same or higher in the same phase. However, the disparity between the teacher's mean score and student's mean score grew in Tasks 2 and 3 of Phase 2, meaning their ratings moved further away rather than closer together when students had the assessment criteria prior to doing the task. Again, this may have been due to the students' perception that Tasks 1 and 2 in Phase 2 were more challenging, which resulted in lower confidence.

Figure 2: Mean scores per task (overall rating on 1–10 scale) for both student and teacher



Another point regarding the overall rating is that there were no clear guidelines for students as to what constituted a 1, 2, 3, 4 etc. The result of this is a consistent lack of relationship between the individual ratings and score out of 10. For example, in Task 1, not one student allocated a score out of 10 that was comparable to the mean score for the five individual ratings (i.e. out of 15). This remained the same for Phase 2, Task 3. It is therefore suggested that this data cannot be relied upon to provide a clear relationship between the use of criteria for self-assessment and an improvement in task performance.

Student perceptions of assessment criteria

During the interviews, all but one student was able to articulate his or her perceptions on the use of assessment criteria and self-assessment. One student was unable to provide coherent answers due to his weak oral skills. His answers were unclear and it was not possible to assess whether he had understood the questions or not. His responses were discounted in this part of the analysis as a result. His self-assessment was still included in the research, however, as his reading and writing skills were deemed strong enough for him to complete the questionnaire and his task performance showed some understanding of the criteria.

When asked if students found it useful to listen to and assess their performance in the speaking tasks, all students said yes. The three main reasons given were:

- you know where your problems are and can improve (e.g. 'I see my fault of speaking or grammar' – Thai student)
- you can note improvement (e.g. 'when you record this week you can see what's you improve' – Arabic student)
- it helps you to use language you can use out of class (e.g. 'sometimes I forget this word when I talk but if I use it in out of school I can use it' – Arabic student).

Half the students said that the assessment form was difficult to complete at first but became easier. When asked whether it was better to have the assessment criteria before they spoke as opposed to after they spoke, it became clear that most students could either not understand the question or were largely unaware that they had completed the self-assessment in two different ways: seeing the criteria before the task and seeing it only after the task. Despite my attempts to paraphrase and explain, students' responses were still largely unclear e.g. 'It's better because we sharing. I know, for example, my friend's problems'. However, one student did say that receiving the criteria in advance of the speaking task was 'sometimes helpful' and one student said that it was 'okay'. When questioned further, both students felt that they already knew what was expected of them without the criteria. This may be because the assessment form was the same throughout the study, with only some information changing to focus on the specific task. Students therefore became familiar with the criteria before they got to Phase 2.

Two students focused on the preparation time given during their interviews. One student found this positive: 'I can preparate [sic] sometimes and I can talk and I can imagine, I can use the grammar, and I can use the past and the present simple. I feel happy with this task or example' (Arabic student). The other believed it was sometimes unnecessary: 'I think the better if we didn't prepare about the subject. If it's about the life now, it's easy to talk about. But if it about technology, yes, we have to prepare' (Arabic student).

In terms of the students' self-assessment skills, five students said that they focused only on the negative points when they assessed themselves, but they found this motivating as it made them want to improve. In fact, six students mentioned feeling motivated by the self-assessment and determined to improve: 'I want to study more' (Chinese student).

Reflections

The results of my research suggest that success criteria may have a positive effect on learning. With regard to my first research question (Does the presentation of assessment criteria **before** a speaking activity help learners to self-assess their speaking skills more accurately?), students did not perceive a difference between receiving the criteria in advance of the speaking task and after. However, there is evidence to suggest that it had a positive impact on the students' ability to self-assess more accurately as their ratings in Phase 2 were closer to that of the teacher. It may be that the act of using criteria to self-assess while listening to a recording was the factor that helped develop their accuracy in Phase 2 rather than when they received the criteria.

In terms of the second research question (If yes, does self-assessment result in an improvement in speaking skills?), students' speaking skills did improve between Phase 1 and Phase 2 according to their individual criterion ratings, with students' performance in the final task in Phase 3 being particularly strong. This may be a result of the criteria provided before they did the task or it may simply be that they had developed their speaking skills as a matter of course through their regular lessons and socialising after school. Having said that, students did feel that self-assessment was useful for them as it allowed them to recognise their weaknesses in speaking, which they could then work on and try to improve.

For me, the overall insights from this research are that the students found recording themselves, listening back and assessing their skills beneficial to their learning and as a result they did become better at self-assessment, regardless of when they had the criteria. Like the students in the Chen study (2008), students underscored themselves in Phase 1 but improved as the study progressed.

The findings from this research have prompted me to continue to regularly encourage my students to use criteria to self-assess their speaking in class. I have also shared the insights with my colleagues at Bell who have experimented with the use of assessment criteria in class for all four skills as a result and reported back some positive results.

The project has also allowed me to understand the process of action research more fully and I value it as both a research tool and a developmental tool. I particularly

enjoyed getting to know my students much better through the interviews and I have since reflected on my perceptions of what my students can and cannot do. I now wonder if I have undervalued their opinions about their own learning in the past and am keen to explore their views on what helps them learn further. My students showed me that, even at a low level, they could articulate ideas on the learning process.

Of course, the research that I conducted was not without limitations. I should have chosen to use simple instructions rather than a complex rubric so the self-assessment questionnaire was easier for students to use at first. If I did similar research again, I would ensure that all tasks are perceived to be of similar complexity. However, as action research is as much about the process as the results, I am pleased with the latter and would recommend action research to other teachers as a means of teacher development.

References

Babaii E, Taghaddomi, S and Pashmforoosh, R (2015) Speaking self-assessment: Mismatches between learners' and teachers' criteria, *Language Testing*, 1–27.

Black, P and Wiliam, D (2001) *Inside the Black Box: Raising Standards Through Classroom Assessment*, available online: weaeducation.typepad.co.uk/files/blackbox-1.pdf

Chen, Y-M (2008) Learning to self-assess oral performance in English: A longitudinal case study, *Language Teaching Research* 12 (2), 235–262.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.

Dörnyei, Z (2007) *Research Methods in Applied Linguistics*, Oxford: Oxford University Press.

Huang, Y and Gui, M (2015) Articulating teachers' expectations afore: Impact of rubrics on Chinese EFL learners' self-assessment and speaking, *Journal of Education and Training Studies* 3 (3), 126–132.

Litz, D (2007) Student-directed Assessment in ESL/EFL: Designing scoring rubrics with students, *The Internet TESL Journal* 8 (11), available online: iteslj.org/Lessons/Litz-StudentDirectedAssessment.html

McMillan, J H and Hearn, J (2008) Student self-assessment: The key to stronger student motivation and higher achievement, *Educational Horizons* 87 (1), 40–49.

Wiliam, D (2011) *Embedded Formative Assessment*, Bloomington: Solution Tree Press.

Appendix: Self-assessment

Listen to your recording and answer the questions. Use the assessment criteria in the box to help you:

1. How well did you describe the person? Tick the correct boxes.

Describe someone in your family	I didn't do it well.		I did it quite well.		I did it very well.	
Task completion (Say what they are like/what is similar/different to you)	I didn't talk about the right things.		I talked about some of the right things.		I talked about all the right things.	
Fluency	I paused a lot.		I paused sometimes.		I didn't pause much.	
Language use (Comparatives/ Superlatives)	I didn't use different structures and words/phrases.		I used a few different structures and words/phrases.		I used several different structures and words/phrases.	
Language accuracy (Comparatives/ Superlatives)	I made a lot of grammar and vocabulary mistakes.		I made some grammar and vocabulary mistakes.		I made only a few grammar and vocabulary mistakes.	
Pronunciation	My pronunciation was difficult to understand. I sounded bored.		My pronunciation was mostly understandable. I sounded quite interesting.		My pronunciation was understandable. I sounded interesting.	

2. Overall, how well did you give your description? Circle a number.

I did the task very badly — **I did the task very well**

1	2	3	4	5	6	7	8	9	10

3. What can you do better next time? How? Please write your answer.

_____

_____

_____

Impact of regular self-assessment on low-level Middle Eastern learners

RICHARD FLYNN ANGLO-EUROPEAN SCHOOL OF ENGLISH, BOURNEMOUTH

CHRISTIAN NEWBY ANGLO-EUROPEAN SCHOOL OF ENGLISH, BOURNEMOUTH

Introduction

'I know it all though.'
'Why are we doing this again?'

These anonymous comments from Middle Eastern learners were frequently heard by teachers in several lessons at our school, Anglo European School of English, a private language school in Bournemouth. After our colleagues informally shared these comments at lunchtime Christian and I began to ask: How aware of their strengths and weaknesses are our learners?

Several feedback questionnaires distributed to all learners upon completion of their course also frequently featured negative feedback from Middle Eastern learners, with complaints about repetition of tasks and objectives made specifically by our Middle Eastern learners from Saudi Arabia, Oman and Libya. During lessons learner dissatisfaction was reported to management by teachers who observed dissipation of learners' initial enthusiasm for classes, reduced concentration levels, uncompleted homework tasks and falling attendance rates.

Dissatisfaction from the Middle Eastern learners was often matched by frustration from their teachers with what they felt was a frequent lack of application from learners. Teachers informally commented upon the substantial difference in their Middle Eastern learners' receptive knowledge and productive output in several writing and speaking tasks. Teachers' repetition of tasks and objectives to develop learners' productive '*performance* in the language' led to conflict with learners who expressed a feeling that they already had demonstrated enough 'knowledge about the language' (Harris 1997:15) to move on to completely different tasks and objectives.

We therefore decided to investigate the influence of cultural and learning backgrounds on Middle Eastern learners' ability to assess themselves and what impact frequent self-assessment would have on low-level Middle Eastern learners enrolled at the school.

The report briefly summarises current thinking on the role and value of self-assessment tasks in the English language classroom before describing the structure of the intervention. We discuss the findings from the two action research cycles before final conclusions on the study and the whole experience itself are drawn.

Teacher versus learner-led assessment

Assessment can be divided into several categories, including summative and formative, with the former 'used mainly to report what has been achieved rather than to help teaching and learning' (Harlen 2006:91) and the latter's function 'is to help the learning of the individual' (Harlen 2006:99). Summative assessment is usually conducted by the teacher at the end of a phase of learning and is used for official measures of learner proficiency or progress. Formative assessment is more focused on identifying strengths and weaknesses, with the aim of addressing specific learner weaknesses.

Summative assessment tends to be more objective; conducted by an 'expert' and used for official grading of learners. In contrast, formative assessment, such as peer and self-assessment, can be conducted and led by learners themselves. Self-assessment gives students the opportunity to reflect objectively, in a non-competitive fashion, on their own accomplishment. Learners can then use their own analyses to focus and improve upon the weakness that they have identified, whilst recognising areas of strength. It is rarely used for formal measures of progress or proficiency. Instead self-assessment seeks to develop learner self-awareness, which is acknowledged as a key element of developing learner autonomy (Thanasoulas 2000).

Self-assessment can be completed in the classroom in a number of ways. The process of self-assessment can be imagined as existing on a line from full teacher to full learner control. Learner-controlled self-assessment may involve learners setting the criteria for assessment or learners choosing what work they are assessing themselves. Reliability, being a measure of how consistent findings are and how results are replicated when using the same measurement procedure (Baumgarten 2010) is an important factor. Indeed, several researchers have emphasised the possible unreliability of self-assessment data as learners lack the capacity to assess their proficiency accurately and frequently 'over-grade' themselves (Khonbi and Sadeghi 2012:58). Evidence suggests 'accuracy is increased when self-assessment is in relation to clear descriptors and related to specific experience' (Council of Europe 2001:191). This could particularly be the case in language learners who are at or near the beginning of their

language learning, as they may lack the linguistic ability to self-assess in the second language. Therefore for this study self-assessment was practised in a more teacher-controlled manner and required the reflection of learners upon a piece of work selected by the teacher and assessment of this work against criteria set by the teacher.

Self-assessment, which requires initial learner reflection, is a key 'pillar' of learner autonomy (Harris 1997:12). Learner autonomy has become accepted by many educational practitioners as a learning goal in itself (Thanasoulas 2000) and this has corresponded with more learner-focused education in mainstream and EFL schools, as seen by the development of the European Language Portfolio (www.coe.int/t/dg4/education/elp/elp-reg/History_ELP/History_EN.asp). In order to become autonomous, learners need to independently and consciously monitor their performances, analyse them and develop a repertoire of efficient learning strategies (Khonbi and Sadeghi 2012:48). Thanasoulas (2000) describes how this greater learner focus reshapes the teacher–learner relationship to transfer more power and authority to the learners themselves. In addition to developing learner autonomy, self-assessment is thought to improve learner motivation by ensuring learner engagement in their own learning processes. This focus on learner autonomy 'does not mean that the teacher becomes redundant' however, and we wanted to investigate what support learners needed to 'develop the skills and mindset that can lead to successful self-guided language study' (Godwin-Jones 2011:4).

Middle Eastern learners and learner autonomy

For Middle Eastern learners, a focus on self-assessment skills is frequently a new experience. Indeed, Baniabdelrahman and Moheidat (2011:56) describe learners' self-assessment in Oman as being 'completely disregarded' by all stakeholders. Many Middle Eastern schools and universities are primarily concerned with the transfer of knowledge from teacher to learner, with the teacher being viewed as the 'absolute authority' in the classroom (Derderian-Aghajanian and Cong 2012:174). Unsurprisingly, therefore, the learners participating in our study had had very limited prior opportunity to direct their own learning and were not previously encouraged to develop learner autonomy or independent learning skills. However, as learners selected for this study had specifically chosen to leave their country of origin to study abroad, it was hoped that they would be more open to alternative methodologies. In order to develop both the motivation and the ability for self-assessment for Middle Eastern learners there needed to be close teacher–learner liaison and discussion throughout. The discussion undertook in the classroom emphasised the purpose of regular self-assessment and continual monitoring from the teacher, and guided learners to devote time and energy towards reflection prior to completing the self-assessment questionnaire.

The intervention

Organisation of the project

The study was an investigation of the impact of weekly self-assessment of set written tasks with low-level Middle Eastern learners. Learners were selected from two levels present at our institution; elementary and pre-intermediate, which roughly approximate to A1/A2 on the CEFR. Learners selected for study were predominantly from Saudi Arabia, with others from Libya, Jordan and Oman. These learners were enrolled for at least four weeks at our school between April and September and were taught in the same class, based on level, as several other learners from different countries. All learners in a class participated in the self-assessment tasks but for the particular focus of the study data presented here pertains only to Middle Eastern learners.

The study involved two cycles of enquiry and each cycle was conducted with a different group of learners. This is because all the learners in Cycle 1 had either left or changed class level by the time Cycle 2 began. Cycle 1 was conducted in April, as the school was not crowded at this time and there were several learners available at elementary level. Cycle 2 was conducted in August, after a busy summer season, when the vast majority of short-term (2-week enrolments or less) learners had left. Due to institutional constraints learners had to be selected from both elementary and pre-intermediate level for Cycle 2. Each cycle lasted for four consecutive weeks to decrease the potential for learners leaving or changing level during the study.

Cycle 1

There were three male participants in Cycle 1, all from Saudi Arabia, with ages ranging from 16 to 24. These learners were studying for a minimum of 15 hours per week and classified as long-term learners. Due to the rolling enrolment policy of the school there were other learners present in their elementary level class, many of whom changed class throughout the period of Cycle 1. Learners had either Christian or me teaching at least one of their lessons per day. We alternated which one of us set and monitored each self-assessment task.

Cycle 1 consisted of initial face-to-face interviews between each participant and one of us, and these interviews were recorded and selectively transcribed. Interviews were followed by four self-assessment questionnaires; each questionnaire was designed to focus on a specific pre-assigned task and was completed in class on a weekly basis under the supervision of their teacher. Each questionnaire, however, was different in style and content (see Appendices 1 and 2) and three tasks were written and one task was spoken. To complete the cycle a written feedback questionnaire was completed by learners at the end of the 4-week-long intervention, again in class.

While these tasks provided a range of data, there was a lack of consistency in that there were different assessment criteria assigned for each task, and the visual style of each form was different for each task. This lack of consistency proved problematic when considering their reliability – which could change according to learners' understanding of the assessment criteria. The main concern regarding the data of Cycle 1 was that the lack of consistency in the self-assessment questionnaires may have significantly impacted upon the findings. It was felt by both researchers that in order to be a rigorous study of the impact of self-assessment, several changes needed to be made for Cycle 2 in the processes used, the questionnaires themselves and in the feedback given to learners during the intervention.

Because we feel Cycle 2 was a more rigorous investigation of the final research question, data from Cycle 2 has been used to establish our findings and Cycle 1 has been seen as a trial intervention which piloted the instruments and processes used in our research. Cycle 1 provided useful feedback for us to plan a standardised assessment criteria questionnaire and suggested our Middle Eastern learners were able to assess their work more accurately over time. It was felt Cycle 2 had to follow on from this first cycle to see if learners could use their reflective skills to actually improve their work.

Cycle 2

For Cycle 2 greater focus was placed on having learners identify and address specific strengths and weaknesses within four language categories – grammar, vocabulary, spelling and punctuation, and task achievement. Learners also addressed self-identified weaknesses by repeating their task. The central research question became: To what extent can low-level Middle Eastern students use self-assessment of written productive tasks to identify their own strengths and weaknesses? To answer this central question we investigated:

a) If low-level learners have the ability to assess their own written production in English reliably and accurately.

b) If Middle Eastern learners are willing to correct their weaknesses autonomously before any teacher corrections.

c) How much scaffolding and support from their teachers low-level learners need for self-assessment.

Cycle 1 informing Cycle 2

Cycle 2 followed the same basic format as Cycle 1, with one-to-one interviews preceding four self-assessment questionnaires and an end-of-intervention feedback questionnaire. Changes implemented for Cycle 2 were based on researcher observations of Cycle 1 and informal feedback received from colleagues, which came mainly from informal discussion in the staff room and focused on the comments the learners had made and any changes the teachers had perceived in the learners' work. The changes made were as follows:

1. Only written productive tasks were assessed by the learners. This simplified the self-assessment process for learners and enabled standardised criteria for each task. Practically, the collection and analysis of learner work was also simpler and the risk of technical issues with recording devices was eliminated.
2. The self-assessment questionnaire criteria (Appendix 3) were standardised to focus on four areas: grammar, vocabulary, spelling and punctuation, and task achievement. Greater standardisation reduced the linguistic demands of the questionnaire for the learners, which was felt would help give more accurate and reliable data.
3. Learners repeated the productive task they had self-assessed so that we could investigate whether learner self-assessment had any impact on improvements in grammar, vocabulary, spelling and punctuation or task achievement. In addition teachers gave individual feedback on the second, final written submission. This change was to encourage learners to complete the second writing task and ensure learners received praise for their efforts, which was important for their motivation.
4. Three questions from the initial semi-structured interview were altered to focus more on learners' ability to describe their strengths and weaknesses as opposed to providing a more general learner discussion of their classes (Appendix 4).

Cycle 2 participants

The three participants for Cycle 2 were all male, from Saudi Arabia and Libya. Data from learner diaries collected by the school on a bi-weekly basis indicated these learners frequently believed they had high productive ability with language covered in class. Counselling sessions conducted by their teacher also highlighted minor learner frustration that they had not advanced in their level swiftly enough. At the time of the study, all learners had been enrolled at the school for at least four weeks and were due to attend a minimum of eight weeks of lessons. The size of the classes in which these learners were enrolled varied throughout the cycle, from four to 14 learners. Detailed learner profiles are given below, with names changed to maintain anonymity.

Muhanned: A 32-year-old learner from Libya. He had been studying English for five months, having started as an absolute beginner. Teachers noted Muhanned's speaking was a particular weakness and the learner admitted lacking confidence in speaking during private learner–teacher discussions. Muhanned had not completed university studies in his own country and intended to study in England for a number of years at an English university. He displayed a positive attitude to learning, with a near faultless attendance record and kept clear vocabulary and grammar notes, which were reviewed at home for several hours daily.

Saad: A 33-year-old learner from Saudi Arabia, who enrolled for a 4-week course during the middle of summer. Saad had completed a university degree in Saudi Arabia

and had visited London for pleasure on a number of occasions. Saad was frequently late for classes and had an attendance rate of 80% over his course of studies. Saad's stated learning aim when enrolling was to improve his general speaking.

Khaled: A 17-year-old learner from Saudi Arabia who was enrolled by his parents in an 8-week course to develop general English skills. In class, teachers noted that Khaled had strong speaking fluency but made repeated spelling errors in writing. It was Khaled's first experience of living and studying abroad; Khaled established no long-term goals for English language learning and frequently failed to complete homework assignments.

Data collection

Data was collected first via written feedback forms and then from spoken interviews. Spoken interviews were semi-structured to provide 'a thematic, topic-centred, biographical or narrative approach . . . but with a fluid and flexible structure' (Edwards and Holland 2013:9), with central questions planned and a list of potential follow-up questions provided to expand upon answers. These semi-structured interviews lasted between 3 and 5 minutes and offered an opportunity to further explain to the participants the nature of the study and the tasks involved.

Written self-assessment questionnaires primarily provided quantitative data as learners marked their assessment of the achievement of task criteria using a 6-point scale. The questionnaire distributed was designed to be as short as possible, covering one side of A4 only. Piloting in other classes highlighted some difficulties the learners had in understanding the self-assessment criteria form, so we spent more time explaining the structure of the form. These questionnaires were completed and collected during the learners' class every Friday afternoon. Each of these classes was taught by Christian or me. The first self-assessment questionnaire and records of the in-class discussion around that self-assessment provided a baseline for the study. Each learner self-assessment was compared to a teacher assessment of the same work against the same criteria. For improved reliability of the teacher assessment both researchers independently marked the tasks before negotiating a final grade, which was checked by a third teacher.

Other data used in the study was primarily qualitative, incorporating records of student–teacher discussions concerning each task. Each self-assessment questionnaire also allowed for open learner suggestions regarding how they could improve their work. The triangulation of qualitative and quantitative data and the range of sources for this data were intended to improve its robustness (Wallace 2008:36). Triangulation of data is said to take place 'when multiple forms of data – when analysed – show similar results, thereby confirming the researcher's findings' (Valcarcel Craig 2009:121).

Ethics

Several ethical issues emerged during both cycles. Of particular concern was whether learners would be unwilling to disclose when they felt uncomfortable given their prior experience of teacher as an 'absolute authority'. To reduce this risk, consent forms offered an opportunity to opt out of the programme and emphasised the possibility of changing their minds at any point in the study. Anonymity was guaranteed to all participants and learners were informed verbally and in writing that we planned to share samples of their work in an article and at a conference prior to each intervention. Several participants took the consent forms home for two or three days to consider whether they would participate and an advanced level Arabic-speaking learner volunteered to explain any elements of the form not understood.

Interviews were conducted during break periods and, in order to reduce disruption to learners, the time and location of the interview was negotiated at least one week in advance and interviews were restricted to a maximum of 10 minutes. An appropriate location for interviews was essential, especially as some participants were under 18 years old. The age of the learners also meant that, in very few cases, permission had to be obtained on their behalf. To reassure the learners we conducted interviews in a quiet, private and well-furnished space. In addition, the potential for learners to be surprised by unexpected questions during the interview was reduced by providing them with a list of topics, in English, to be covered prior to the interview.

The intrusiveness of the weekly questionnaire was considered carefully, especially as it was completed during learners' class time. To reduce the intrusiveness a time limit was established and a fixed time and day was set for self-assessment. The day chosen was already an established part of our school's Friday routine focusing on self-reflection tasks and individual counselling sessions.

Main findings

As both teachers felt that the data from Cycle 1 was too unreliable, the findings presented here are based solely on the data collected from Cycle 2.

First, learners began to improve the accuracy of their self-assessment at the end of the 4-week period (see Figure 1). The least number of differences in learner and teacher assessment were to be found in assessing vocabulary, for which the most specific criteria and clearest models of target language were given. In Figure 2, the percentages show the differences between the learners' self-assessments and the teachers' assessments in each category. These are expressed as percentages as the learners and teachers could find multiple errors in each category.

Learner comments regarding ways to improve their productive task became more task specific after the first week. Initially learners could or would not identify ways to improve their work, with two of the three learners

Figure 1: Number of differences between self and teacher assessments

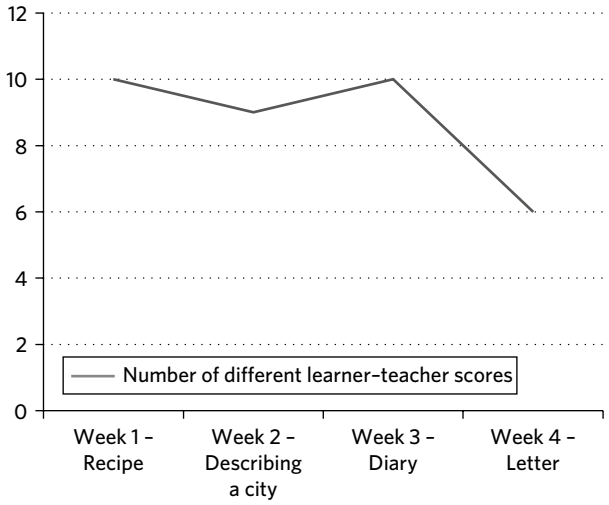
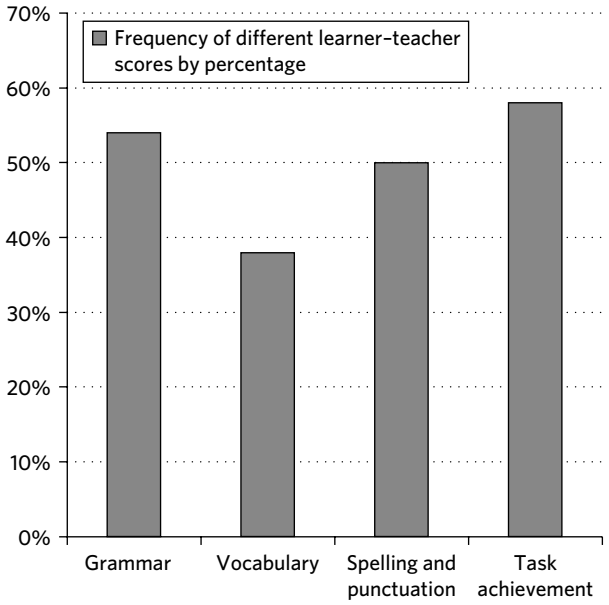


Figure 2: Differences in learner–teacher assessment by category



writing 'nothing' when asked how to improve their work. After the first week, however, all learners displayed an ability to identify one area to focus their improvement on. Muhanned in particular was able to identify key vocabulary or grammar changes that could be made. Saad and Khaled's suggestions on how to improve were quite erratic, varying from the specific vocabulary-focused 'write more names for places' to the more general 'use more grammar and vocabulary'. All learners however could identify specific ways to improve their work on at least one of the four tasks (highlighted in bold in Table 1).

In addition, during Week 3 and Week 4 Muhanned asked several specific questions about his work during Friday's self-assessment class time and in Week 4 he independently underlined parts of his productive task that he felt had errors, before requesting teacher assistance to advise on these errors.

Table 1: Learner comments on ways to improve their productive task

	Week 1 – Recipe	Week 2 – Describing a city	Week 3 - Diary	Week 4 – Letter
Muhanned	Nothing	**Use could, must, have to**	**Change words for adjectives (really, very, so, incredible)**	**Practise look at, for, of, to**
Saad	Nothing	**Write more names for places**	Use more new words	Use more grammar and vocabulary
Khaled	Other words	**Use could, must, should, have to, to give ideas**	**Other words, I'll use more next, then, after**	Look carefully for mistakes

The number of occasions where learners completed and submitted a second attempt at their productive task was very low. Overall the productive task was repeated on only five occasions in total, out of a maximum of 12. It is our belief that, for the majority of the participants, the task was viewed as 'finished' and there was no need to repeat the task.

However, there were several individual differences in the learners: Muhanned repeated the set task on three out of four occasions, whilst Saad and Khaled only repeated the productive task on the first of four tasks, and when they did so they made very few changes. Muhanned displayed the highest number of changes between his first and second attempts at each productive task. Sample changes made to Muhanned's description of his city task are shown below and additional information given is highlighted in bold:

> Attempt 1: *The city has Mediterranean climate. The climate's cosy and wonderful.*
>
> Attempt 2: *The city has Mediterranean climate* ***so, if you visit that place in the summer, you should wear summer clothes and hat or cap because the weather's very hot****.*

One of the changes made by Muhanned, the use of 'should', had been highlighted as an area to improve on his self-assessment questionnaire, indicating his conscious use of his own self-assessment feedback during his second attempt at the productive task.

Motivation to self-assess varied significantly between the three learners and for Saad and Muhanned; their levels of motivation for self-assessment did not appear to change over the four weeks. At the end of Cycle 2, Saad expressed a strong desire not to continue to self-assess to his teacher while Muhanned actively sought to continue the programme. Khaled, in his original interviews, had displayed a dislike for editing and assessing his own work but by the completion of Week 4 also requested to continue the programme.

Key issues

The accuracy of learner feedback and its gradual improvement over the course of the study was affected by several factors. One key issue was learner difficulty with the initial understanding of the self-assessment process. Learners found it challenging in the first and second weeks to focus their responses on one particular task. Instead they frequently said their grammar was poor, for example, and supported this assertion by referencing their general class work, not the specific productive work being assessed. Therefore, learner feedback was initially not always task specific and their responses may have been based more strongly on general teacher feedback.

However, after completing two feedback questionnaires learners were more able to focus on assessing the selected piece of work only. They spent more time reading through the specific productive work itself and they answered the more open question of how they could improve their task by referring specifically to the work being assessed.

Whilst no quantitative data were gathered on the length of time taken by learners to look at their work, from our observations it appeared that as learners became used to the process they would spend more time reading their own work before completing the self-assessment questionnaire. We feel that the improved accuracy of learners' self-assessment may have been partially the result of an increased length of time used for reflecting on their work.

The vocabulary category may have been the most accurately assessed as learners were provided with clear models of target language use and therefore the task of searching for these words in their productive work was the least cognitively challenging. Assessment of task achievement however is more subjective, which may account for teacher and learner differences when assessing this area.

Learners' ability to identify key language that could improve their work seemed to develop significantly after Week 1. This change after Week 1 may have been because learners realised that they would have to complete this self-assessment before they received any teacher feedback or corrections. As learners knew it would take longer to obtain teacher corrections they possibly focused on more immediate corrections they could make themselves.

More specific learner suggestions on how to improve their work may also have been assisted by the establishment of fixed periods in class devoted to private self-reflection and assessment. Learners knew that they would have time and opportunity to discuss their own work with their teacher and possibly felt more comfortable admitting weaknesses after regular and personal teacher–learner discussions. These more private and personalised discussions showed learners that their own reflections and assessments were appreciated, valuable and would be acted upon by teachers.

Learner willingness to repeat their task varied significantly between learners. Muhanned, who had expressed a keen desire to improve his English to study at a UK university, demonstrated the greatest willingness to repeat a task. Additionally, learners who had written their learning targets were more likely to repeat the task; learners focused on speaking may not have seen the relevance of repeating a writing task. The overall aims of each learner therefore impacted significantly upon their attitudes towards attempting specific tasks twice.

Reluctance to repeat a task may also be the result of more confident learners feeling that they had already fully achieved the task; for example, very few learners repeated the first task of writing a recipe, which was also the task all three learners assessed as being 'very easy'. Alternatively, in the case of Saad, his lack of clear and organised language notes may suggest learners lacking clear examples of lexis or grammar struggles to address self-identified weaknesses.

Limitations

There are clear limitations with the accuracy of data gathered from learners themselves. Even with more objective teacher-set criteria, learners' self-assessment is subjective for each individual and their responses could be affected by their varying emotions, moods, confidence levels and degrees of motivation. As acknowledged by Wallace (2008:127), when asking learners to answer questions about their own learning and application, as researchers we have 'no way of ascertaining the truth of the reply'. However, attempts to mitigate these limitations were made by our monitoring of learners and one-to-one support throughout the intervention.

In addition, as learners were all at low levels of English ability, linguistic confusion could have impacted upon the data received and more open questions requiring learners to provide comments on how they would improve their work would be limited by their ability to express themselves in English. It is expected that there would be some margin for miscommunication in this section as our learners may lack the language to explain their ideas clearly. However, teachers monitored learners' self-assessment to verbally check their ideas and reduce the potential for miscommunication.

As the study was conducted over a relatively short period of only four weeks and, with a small sample of learners, the applicability of our findings to a wider cohort of learners in our school cannot be measured. The learners we worked with had very similar learner profiles to many other low-level Middle Eastern learners that had enrolled in our school in previous years and the variation in their English language targets, goals and motivations suggest that the participants actually covered a fairly broad spectrum of low-level Middle Eastern learners.

Practical value

Our learners displayed greater accuracy over time in highlighting their own weaknesses or difficulties when assessing their productive task in that

their self-assessments more closely matched their teachers' assessments. Self-assessment was a fixed, once-a-week activity and most learners displayed willingness to complete the initial task and weekly self-assessment, especially when they knew they would be able to spend time discussing their work with their teacher. This suggests that an element of reward – here in the form of learner–teacher discussion – was important to learner motivation as one-to-one time with the teacher was seen as very valuable by these learners.

While self-assessment is frequently described as being one element of developing learner autonomy our research suggests that this development could be used by teachers to guide the learners to identify weak language areas and to set specific language aims, thereby making the learners more autonomous. As learners highlighted errors and language difficulties not directly assessed in the questionnaire, the teacher and learner could use self-assessment to organise a more personalised syllabus. In fact as a result of our research our school began to tailor short one–to-one sessions to address specific problems selected by low-level Middle Eastern learners themselves. This took the form of a 'student surgery'. The students could book a 15-minute slot with the Director of Studies, Assistant Director of Studies or the senior teacher in which to be given guidance on a specific language point which had been previously chosen by the learner.

One possible advantage of learner self-identification of strengths and weaknesses is that as the weakness is self-identified some learners may be more motivated to address this weakness. On occasion learners and teachers can have conflicting opinions over success in a task and language errors, and some learners may resent having their weaknesses pointed out to them by others. But when learners found weaknesses themselves they appeared more accepting of them and there seemed to be a less negative effect on their confidence.

The practice of self-assessment, with teacher-directed criteria, also increased learner–teacher dialogue in the classroom. This helped to mitigate conflict between the two parties, as it provided an opportunity for discussion and negotiation of language aims. A change of mindset, evidenced by the decrease in differences between self-assessments and teacher assessments, for Middle Eastern learners, from viewing learning as something done to them rather than something they do themselves required a change in the teacher–learner dynamic, from the teacher being the authority to being a counsellor. To try and achieve this change it appeared that personalised feedback and private one-to-one time and discussion was important, as it established a more equal relationship between learner and teacher and provided an opportunity for learners to take greater ownership of their learning. Connected to this point, our research supported views that becoming an autonomous learner is an ongoing process that requires both substantial time on a regular basis in the classroom and constant teacher support and guidance.

Reflections

As can be seen in our research, there were significant differences between Cycle 1 and Cycle 2 in the instruments used and data collected. This shows that it is vital to use experimentation, piloting and the first cycle to inform the direction of study. For future action research participants, we would suggest a degree of flexibility to be inbuilt into plans in order to adapt to a constantly changing research situation and to cover unexpected events. We would also suggest utilising all resources available, colleagues especially, as they can pilot tasks, offer impartial feedback and provide informal observations of the learners involved in the project.

Personally we have found the action research project challenging, demanding and engaging. We both feel we have learned ways to reflect and improve on our teaching and feel confident in continuing smaller scale action research programmes with learners. From this study we will ensure that attempts at learner autonomy are continuously supported by teacher input. Without any teacher guidance our learners clearly became disengaged with self-assessment and self-reflection. We plan to continue a focus on self-assessment in lessons and would like to investigate developing self-assessment skills in connection to learning portfolios. We both believe a similar intervention to the one displayed here could benefit learners who struggle to change to classes at another level – involving them in their own learning assessment and establishing goals together could help them to make progress again. It would also be interesting to investigate whether learners could transfer their reflective skills and self-assessment ability to other tasks. Our study and the data gathered were limited by the teachers actively involved in the research; to see whether learners were more able to self-assess other tasks, their speaking for example, would have required much wider teacher participation.

References

Baniabdelrahman, A and Moheidat, A (2011) The impact of Omani 12th grade students' self-assessment on their performance in reading in English, *Asian EFL Journal* 13 (1), 48–84.

Baumgarten, M (2010) *Paradigm Wars – Validity and Reliability in Qualitative Research*, Munich: GRIN Verlag.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.

Derderian-Aghajanian, A and Cong, W (2012) How culture affects on English language learners' outcomes with Chinese and Middle Eastern immigrant students, *International Journal of Business and Social Science* 3 (5), 174–180.

Edwards, R and Holland, J (2013) *What is Qualitative Interviewing?* London: Bloomsbury Academic.

Godwin-Jones, R (2011) Emerging technologies: Autonomous language learning, *Language Learning and Technology* 15 (3), 4–11.

Harlen, W (2006) *Assessment and Learning*, London: Sage.

Harris, M (1997) Self-assessment of language learning in formal settings, *ELT Journal* 51 (1), 12–20.

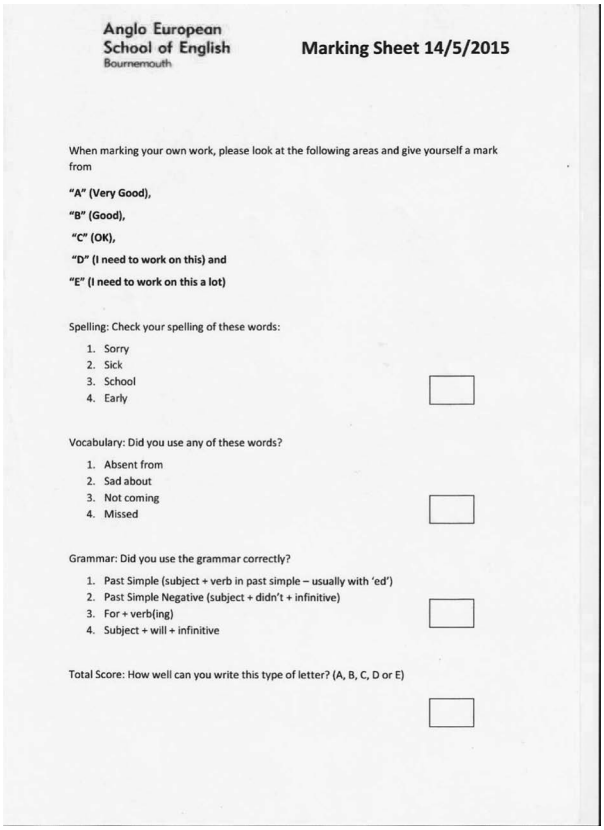
Khonbi, Z and Sadeghi, K (2012) Effect of assessment type (self vs. peer vs. teacher) on Iranian University students' course achievement, *Language Testing in Asia* 2 (4), 47–74.

Thanasoulas, D (2000) What is learner autonomy and how can it be fostered? *The Internet TESL Journal* 6 (11), available online: iteslj.org/Articles/Thanasoulas-Autonomy.html

Valcarcel Craig, D (2009) *Action Research Essentials*, New York: John Wiley & Sons.

Wallace, M (2008) *Action Research for Language Teachers*, Cambridge: Cambridge University Press.

Appendix 1: First example of a self-assessment questionnaire for written task in Cycle 1



Anglo European School of English
Bournemouth

Marking Sheet 14/5/2015

When marking your own work, please look at the following areas and give yourself a mark from

"A" (Very Good),
"B" (Good),
"C" (OK),
"D" (I need to work on this) and
"E" (I need to work on this a lot)

Spelling: Check your spelling of these words:

1. Sorry
2. Sick
3. School
4. Early

Vocabulary: Did you use any of these words?

1. Absent from
2. Sad about
3. Not coming
4. Missed

Grammar: Did you use the grammar correctly?

1. Past Simple (subject + verb in past simple – usually with 'ed')
2. Past Simple Negative (subject + didn't + infinitive)
3. For + verb(ing)
4. Subject + will + infinitive

Total Score: How well can you write this type of letter? (A, B, C, D or E)

Appendix 2: Second example of self-assessment questionnaire for written task in Cycle 1

My writing: Write a dialogue with a customer and a shop assistant in a clothes shop.

Example:

Shop assistant: Hello. *Can I help you*?
Customer: Yes, I'm looking for some blue jeans. *Do you have any*?
Shop assistant: Yes, that's no problem, *what size would you like*?
Customer: 32, please.
Shop assistant: No problem. *Would you like to* try these on?
Customer: Yes, please. *Where are the* fitting rooms, *please*?
Shop assistant: Just over there.
(Some time later)
Customer: They look great. *How much are these jeans*?
Shop assistant: *They're* £30 and this red t-shirt *is* £10. I think the medium will really suit you.
Customer: Ok, I'll take them.
Shop assistant: Here you are.

Read the sentences below and tick the sentence if it is true for your writing.

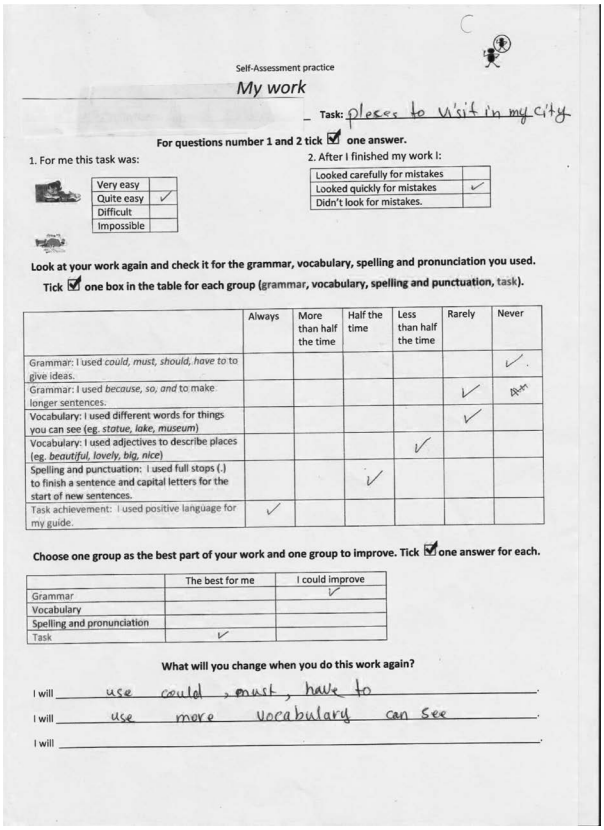
1. I used different adjectives to describe clothes.
2. I used some of these expressions from a clothes shop: 'I'm looking for. . .', 'try on', 'fitting rooms', 'I would like', 'go with', 'suit you', 'here you are'.
3. *I used the right words in the right order for my questions.*
4. *I used are for plural objects and is for single objects.*
5. *I used some long sentences with more than 3 or 4 words in them.*
6. I used the right spelling for most of the words.
7. **I used question marks (?), commas (,) and full stops (.) in the right place.**

Think about your vocabulary (sentences 1 and 2), *grammar (sentences 3, 4 and 5)*, spelling (sentence 6) and **punctuation (sentence 7)**.

What do you think was good about your writing?

What do you think you could improve in your writing?

Appendix 3: Example of Cycle 2 self-assessment questionnaire, both of which followed the same format and structure



Self-Assessment practice

My work

Task: please to visit in my city

For questions number 1 and 2 tick ☑ one answer.

1. For me this task was:

Very easy	
Quite easy	✓
Difficult	
Impossible	

2. After I finished my work I:

Looked carefully for mistakes	
Looked quickly for mistakes	✓
Didn't look for mistakes.	

Look at your work again and check it for the grammar, vocabulary, spelling and pronunciation you used.

Tick ☑ one box in the table for each group (grammar, vocabulary, spelling and punctuation, task).

	Always	More than half the time	Half the time	Less than half the time	Rarely	Never
Grammar: I used *could, must, should, have to* to give ideas.						✓
Grammar: I used *because, so, and* to make longer sentences.					✓	
Vocabulary: I used different words for things you can see (eg. *statue, lake, museum*)					✓	
Vocabulary: I used adjectives to describe places (eg. *beautiful, lovely, big, nice*)				✓		
Spelling and punctuation: I used full stops (.) to finish a sentence and capital letters for the start of new sentences.			✓			
Task achievement: I used positive language for my guide.	✓					

Choose one group as the best part of your work and one group to improve. Tick ☑ one answer for each.

	The best for me	I could improve
Grammar		✓
Vocabulary		
Spelling and pronunciation		
Task	✓	

What will you change when you do this work again?

I will use could, must, have to.

I will use more vocabulary can see.

I will

Appendix 4: Interview questions and prompts given to interviewer for one-to-one learner interviews before Cycle 2

Interviewer welcomes student and asks if the sound from the interview can be recorded.

Student is told that the interviewer is to ask questions about how the student feels about lessons, and the strengths and weaknesses in the written and spoken work they produce in class.

Before the interview is recorded students are asked to bring in a piece of work they think shows their language strengths, and before recording of the interview begins, the questions are distributed to the student and any vocabulary difficulties or student anxieties about any of the questions are discussed.

Name of student:
Date and time:
Location:

1. How are you feeling today?
2. How do you feel about your lessons this week? (*What did you enjoy about them? How could they be better? Which one was your favourite? Why?*)
3. What work have you selected to show and why? (*Did you enjoy the task? How did your teacher respond to this work? Can you show me what the best part of the work is for you?*)
4. Do you think you could improve this work? (*How? If you did it again now how would you change it?*)
5. What are your strengths and weaknesses in English? (*Reading? Writing? Grammar? Listening? Speaking? How do you know these are your strengths/weaknesses?*)
6. How do you feel assessing your own work? (*Do you think you can do it on your own? Do you think it's better to do with the teacher's help or on your own?*)
7. How often do you assess your own work in class? At home? (*Could you do it more? Do you prefer to do it in class or at home?*)
8. How would you feel assessing one piece of work every week? (*Do you want to choose the work or do you think the teacher should? Would you prefer to focus mainly on the strengths/weaknesses/both?*)

The effects of a productive skills focus on summative assessment

APRIL PUGH CELTIC ENGLISH ACADEMY, CARDIFF

CERI THOMAS CELTIC ENGLISH ACADEMY, CARDIFF

Introduction

At our institution, which is a private language school for students visiting the UK to improve their level of English, we require our students to sit weekly tests in order to assess the learning which has taken place during that week. The weekly tests that are given to our students typically consist of discrete items which test their passive knowledge of grammar and vocabulary. To us, this seemed to be at odds with the tests which we give our students at half-term and at the end-of-term, as these tests require students to show productive language use through writing and speaking tasks. We thus felt that our weekly tests did not adequately prepare them for the mid-term and end-of-term tests. Given the importance of the latter in particular in deciding whether or not a student should progress from one level to the next, we felt it was an area which would be worthy of investigation.

Another motivation for our research was that a large number of our students simultaneously take an *IELTS* test and we felt that the weekly tests did not really help these students to prepare for the *IELTS* test, which requires students to use English productively. Given that most of our students study 21 hours a week, we were also keen to investigate whether spending up to 1.5 hours (which represents approximately 7% of weekly class time) on testing was an effective use of time. Therefore, the aim of our project was to better understand whether or not a greater focus on productive skills in weekly tests would help students to achieve improve their performance in the speaking and writing tasks in the mid-term and end-of-term tests.

Literature review

The benefits and drawbacks of testing

Our institution's weekly progress tests are used as a means of formative assessment and tend to focus heavily on discrete item testing. Summative assessment takes place at mid-term and the end-of-term and these tests consist of a mix of discrete test items and productive skills testing. As noted above, one of the concerns we had was whether or not weekly testing is a beneficial use of class time. This issue is also raised by Harris and McCann (1994), who state that a common complaint made by teachers about testing is that it takes up time which could be better used for teaching. In addition they cite three other common complaints about testing which are:

- the teacher already knows what their students are capable of
- tests may not reflect how hard individual students have worked
- tests are demotivating for students.

However, Harris and McCann (1994:26) argue that tests 'should be seen as part of the teaching/learning process and not divorced from it . . . [and] enable us to measure progress in a more individualised way' and encourage students to 'think about their problems and do something about them'.

Washback

Another consideration about testing is the effect which it has on teaching. This is known as washback and according to Hughes (2003:1) it can be 'harmful or beneficial'. In terms of harmful washback, Baxter gives the example of a fluency-based syllabus which is assessed by means of a multiple-choice grammar and vocabulary test leading the teachers to 'teach grammar and vocabulary rather than fluency' (1997:28). According to Harris and McCann, positive washback can be achieved through tests which reflect 'what we are doing in class' and that 'if tasks assess communicative ability, our students will be encouraged to take part in communicative activities' (1994:34).

Formative and summative assessment

Baxter draws a distinction between summative evaluation and formative evaluation. The former is done 'during a process' whereas the latter is 'a kind of final assessment, summarising what has been achieved during the course' (1997:32). At our institution, a course lasts for 12 weeks, and students are given a progress test each Friday as a means of formative assessment. According to Lambert and Lines (2000), weekly tests, which may be categorised as formative assessment, provide feedback to students and teachers to support future learning. This view is also put forward by Baxter, who states that formative assessment allows teachers 'to check how successful the teaching programme is' (Baxter 1997:32). Harris and McCann state that these types of tests can give information to the teacher as well as 'provide important feedback to the student' (1994:28). At our institution summative assessment takes place at the mid-term and end-of-term points and the students' performance in the latter is the deciding factor

in whether or not a student should proceed to the next level. Harris and McCann make the point that summative assessment can 'put a lot of stress on both teachers and students' (1994:28).

Discrete item testing

The method of formative assessment we use in our school is based on discrete items to check understanding of grammar and vocabulary. According to Baxter (1997), there are six usual types of exercise which are used in discrete item testing:

1. True/false questions
2. Multiplechoice questions
3. Gap-fill exercises
4. Editing
5. Sentence building
6. Mistake spotting

The benefits and limitations of discrete item tests

Discrete item tests have certain advantages, including being 'easy to mark' (Baxter 1997:35) and providing 'high reliability' (Harris and McCann 1994:34). Whilst discrete item tests carry some advantages, they also have some disadvantages. First of all, as Baxter (1997:35) points out, multiple-choice questions and true/false questions 'do not test the student's ability to produce or use correct language' as they simply test recognition. This view is echoed by Harris and McCann, who state that discrete item tests 'can have low validity because doing a test such as multiple choice is not a test of real communication' (1994:34). Furthermore, Hughes states that: 'If we want to know how well someone can write, there is no way we can get a really accurate measure of their ability by means of a multiple-choice test . . . We may be able to get an approximate measure, but that is all' (2003:3).

Hughes also states that: 'students' true abilities are not always reflected in the test scores they obtain' (2003:3). In the past we have noted this to be the case with the quieter students who come from a culture of testing and excel in weekly discrete item tests, but who are unable to make productive use of the language. Whitehead and Manassian believe that testing should go beyond passive understanding and state that testing tasks 'need to make sure that task focus is on what the student can do more than on what the student knows about language' (2014:187). A final disadvantage of this type of test, especially multiple-choice items, is that students might simply guess the right answer.

Productive skills testing

Discrete item tests are based on the belief that 'language can be broken down into its component parts and that those parts can be tested successfully' (Douglas-Brown 2004:8). This belief, however, is at odds with the unitary trait hypothesis and the principles of communicative language testing. Although many language tests assess the four skills separately, according to the unitary trait hypothesis, 'vocabulary, grammar, phonology, the "four skills" and other discrete items cannot be disentangled from each other in language performance' (Douglas-Brown 2004:9). The latter, which started to gain ground in the 1980s, is based on the belief that there needs to be 'a correspondence between language test performance and language use' (Douglas-Brown 2004:10).

Benefits of productive skills tests

Whereas discrete tests are a form of indirect testing, speaking and writing tests are a form of direct test and can carry certain benefits. In terms of writing, Heaton notes that 'composition writing can be a useful testing tool. It provides students with an opportunity to demonstrate their ability to organise language material, using their own words and ideas, and to communicate' (1988:137). Whilst oral exams may be a source of anxiety for some students, Lambert and Lines state that oral assessment can be 'very motivating to some students' (2000:149). Given our interest in the impact which a greater focus on productive skills in formative testing may have on students' performance in summative assessment, we were particularly interested in a study carried out by Gong (2010), which took the form of a 1-year comparative study of a control group and an experimental group. This study focused on the adoption of a task-based approach to testing third year college students, who were majoring in English, in Taiwan. Class 1 (the control group) used its usual syllabus, teaching method and assessment approach. The assessment approach for the control group was an indirect test of spoken English and comprised exercises such as sentence reading (in other words, reading aloud). In the experimental group, class 2, a task-based interview was used as a means of assessment and a relevant syllabus was designed. At the end of the year, the two groups were given a mock-up *IELTS* Speaking test. Although the students were assessed against the four *IELTS* speaking criteria of Fluency and Coherence, Lexical Resource, Grammatical Range and Accuracy and Pronunciation, a 100-point marking scheme was used rather than the 9-band marking scale usually used in *IELTS*. In this test, it was found that 'the experimental class performed much better than the control class' (2010:4) with the control group achieving a mean score of 77 and the experimental group achieving one of 86.

We were also interested in a study carried out by Salvisberg (2014) which involved two classes of university students who were taking a course in Human Resource Management. The action research was based on organising a mock exam to see if it produced improvement in the students' performance in the final exam. The results of the study showed a slight improvement in the students' oral skills and revealed that the students were overwhelmingly in favour of the mock exam. The report concluded that 'the end does very definitely justify the means' (Salvisberg 2014:184). We felt this study suggested that a greater focus on productive skills could bring about benefits to the students.

The drawbacks of productive skills tests

Despite the advantages of productive tests, they also carry some disadvantages. Heaton points out that 'the chief objection to the inclusion of the composition question as part of any test is generally on grounds of unreliability' (1988:144), in other words, marking writing can be more subjective than marking a discrete item test. He goes on to state that 'one effective way of increasing test reliability in such cases is by means of a carefully drawn up banding system or rating scale' (1988:165). Harris and McCann note that subjective tests 'can be objectivised by using rating scales which outline a description of what each point on a scale means' (1994:55) and Heaton also states that 'banding systems devised for a particular group of students at a particular level are far more preferable to scales drawn up for proficiency tests administered on a national or an international basis' (1988:145). In terms of oral assessment Lambert and Lines point out that this can be very 'time-consuming' (2000:149).

In brief, assessment by means of discrete item testing has the benefits of being easy to mark and reliable, but it tends to focus on passive understanding of language rather than revealing what the students are able to do with language. Productive skills testing provides students with the opportunity to display what they can do with language and can be more motivating for students, but it can be more unreliable and more subjective.

Our study

Burns notes that one of the main aims of action research is 'to identify a problematic situation . . . that the participants consider worth looking into more deeply and systematically' and 'to intervene . . . in order to bring about changes and, even better, improvements in practice' (2010:2). One of the first steps in action research is to define a research question. Anderson points out the importance of formulating a question with 'just the right scope' (1998:38). Given our interest in the lack of focus on productive skills in weekly tests and the disparity between these and the mid-term and end-of-term tests, we decided to base our action research project on the following question as we felt it was achievable, realistic and measurable within the time frame which we had available: To what extent, if any, does a focus on productive skills have on the students' performance in writing and speaking tasks in the mid-term and end-of-term exams?

The context

The research took place at Celtic English Academy, which is a private English language school located in Cardiff. The majority of our students come from the Middle East (53%), predominately Saudi Arabia and Iraq; Northern Africa, mainly Libya; and South East Asia, mainly Japan and South Korea. During the summer, we usually have quite a high percentage of European students but their number drops significantly from 46% in September to 20% in June. Students from the Middle East and Libya tend to have aspirations to study at university in the UK and many of them take an *IELTS* test during their time with us. European students tend to study for shorter periods of time, usually between one week and three months, though a small minority study for longer than this. Students from the Middle East and South East Asia tend to be long-term students and typically stay at the school for three to nine months. The school has a maximum class size of 10 and the average class size is around eight. Students usually study for 21 hours a week, although a small minority study for 15 hours or 30 hours a week.

We decided to use two groups for our study: a control group and an intervention group, both of which were studying at pre-intermediate level and most of whom had been studying at the school for some time and were therefore used to the weekly testing system. We felt that this level would be able to cope with the demands of extended speaking and writing tasks and at a good level to map progress. Although the two groups were studying different coursebooks, which may have affected their comparability, they were following the same syllabus. The control group was taught by Ceri Thomas and the intervention group was taught by April Pugh. Timetable constraints and student numbers meant that we were unable to have a similar nationality mix in both groups. The control group consisted mainly of students from South East Asia, whereas the intervention group consisted of students mainly from the Middle East.

The planning stage

Some initial decisions

Before the study commenced we made the following decisions:

1. The students in both classes would receive an initial writing and speaking test, which would provide us with baseline data about their writing and speaking skills. According to Hendricks, baseline data is important as it allows comparisons to be made 'before and after the intervention occurs' (2006:97). He goes on to say that only measuring students' writing after the intervention has occurred would make it impossible to measure the effect of the intervention on the students' writing.
2. We decided that the intervention group would receive a productive test each week (with a focus on either speaking or writing) which we would design ourselves, whereas the control group would take a standard weekly test which consisted of discrete vocabulary and grammar test items taken from the test which accompanied the coursebooks we were using.
3. In both the mid-term and end-of-term tests, both groups would be given the same writing and speaking tasks in order to evaluate what difference, if any, the greater focus on productive skills in the weekly tests had had

on their performance in the mid-term and end-of-term tests.

4. As we were keen to focus on the effect that testing had on their productive skills, we agreed that we would not make any changes to our teaching methods and that the control group and the intervention group would continue to perform speaking tasks in class and writing tasks both in class and as homework (but not as part of their weekly test). Therefore the only difference between the two groups in terms of our intervention was the nature of the weekly tests which the experimental group received.

Data collection methods

At the planning stage we collected data as follows.

Results of the students' tests

These provided us with quantitative data and allow us to measure students' performance in both classes at several points in the study. The importance of marking scales in speaking and writing has already been noted and we therefore created marking scales which could be used and adapted during the course. For example, in the writing test we devised a writing scale based on the criteria of grammatical range, grammatical accuracy, lexical range, lexical accuracy, task completion, spelling, punctuation, coherence and cohesion.

Field notes

Field notes are a way of 'reporting observations, reactions and reflections' (Hopkins 2008:104) by keeping a written record of what happens in class. As they 'can reflect general impressions of the classroom' (Hopkins 2008:105), we felt they would be useful to monitor how students reacted to tests and whether there was any difference in their reaction to working productively in class time compared to their reactions to working productively in tests. We also felt that it would be useful to compare how students, especially those in the control group, performed when speaking in class.

Questionnaires

Whilst the main focus of our project was on students' performance in tests, we were curious to ascertain their thoughts about testing. We therefore decided that we would use questionnaires at the start of the study and then again after the mid-term and end-of-term tests to ascertain their thoughts. According to Tisdall, Davis and Gallagher, questionnaires should be kept 'as short as possible, written simply and clearly' (2009:74). We therefore decided to keep our questionnaires short and to the point and to support the design of our questionnaires with pictures to represent the different responses, because as Tisdall et al also point out, 'graphics can be useful' (2009:74) and we felt that the students would be better able to understand the meaning of the possible responses.

Focus group

Related to our desire to find out what the students felt about testing we wanted to supplement the findings of our questionnaires with more qualitative data provided from the students themselves. Whilst interviews can be a 'rich source of data' they are also 'very time-consuming' (Hopkins 2008:110). Another disadvantage of individual interviews is that 'many individuals feel pressure to want to give answers to interview questions that please the interviewer and many try to guess the answer favoured by the person asking the question' (Campbell, McNamara and Gilroy 2004:99). Whilst one of the dangers of focus groups is that some students might not want to give their opinions in front of others and some students may dominate, we felt that the students would have greater confidence to express their feelings rather than please the questioner if they were part of a group. We therefore opted to carry out a focus group session following the mid-term and end-of-term tests.

Ethical considerations

At the planning stage we also considered the ethics of our project. The need for this is pointed out by Borg (2010:10) who states that 'good quality research is ethical' and we were mindful that we needed to make sure that all the participants were aware of the study. Burns highlights the need for the informed consent of the participants, stating that they 'have a right to know about the purpose, the procedures, possible effects of the procedures on them, and how the research will be used' (2010:35). We decided that we would discuss the project with our students and before the intervention commenced we designed a consent form to make sure that the students were happy to take part in the study.

To summarise, after formulating and refining our research question we planned how we were going to undertake our research and considered the four main tools which would allow us to gather the necessary data, which was the results of the students' tests, field notes, questionnaires and focus groups. We also considered the ethics of our project. Once we had considered these factors, we were able to design the tools which we needed to commence Cycle 1.

Cycle 1

The students

Cycle 1 lasted for six weeks and involved the two classes previously mentioned. The intervention group consisted of six students, all of whom were Arabic speakers. The control group consisted of four students: two from South Korea, one from Japan and one from Italy. As the school operates a policy of continuous enrolment, there were at times other students in each class. For example, in the control group there was an Italian student and a Turkish student at the start of the course, but these students left before the mid-term test. As a result we have excluded the results of short-term students from our results. The experimental group was taught and assessed by April Pugh and the control group was taught and assessed by Ceri Thomas.

The first day: What we did

Once the students had been made fully aware of the project and their consent gained, the students completed:

1. An initial questionnaire which provided data about their attitudes to language testing (Appendix 1). In addition, they were asked to comment on what they liked and disliked about tests to provide some qualitative data.
2. An initial writing test (Appendix 2) which provided baseline data about their writing skills. The test comprised two writing tasks. The first task required students to give a description of a place they had visited and the second task required students to give a description of their family.
3. An initial speaking test (Appendix 3) which provided baseline data about their speaking skills. This consisted of two speaking tasks. The first consisted of a question and answer speaking test, requiring the students to ask some basic questions about themselves and the second was a long turn monologue requiring the students to describe a person who they knew.

The first day: Baseline data

The results of the initial writing test and speaking test as well as the questionnaire are presented in Table 1.

Table 1: Baseline writing and speaking marks

Group	Average initial writing mark out of a total 40	Average initial speaking mark out of an initial 40
Control group	18	12.4
Intervention group	17.2	13.8

Table 1 represents the average mark achieved by both groups in the initial writing and speaking tests. It shows that both groups had roughly the same skills in writing at the start of the course, with the control group's writing skills being slightly higher than those of the intervention group. For the control group the lowest mark in the initial writing was 14 out of 40 and the highest mark achieved was 25 out of 40. The range of marks was lower in the intervention group, with the lowest mark being 11.5 and the highest 22.

It also shows that the intervention group had slightly higher speaking skills at the start of the course. For the control group the lowest mark in the initial speaking test was 8 out of 40 and the highest mark achieved was 15 out of 40. The range of marks was less marked in the intervention group with the lowest mark achieved being 11 out of 40 and the highest mark being the same as that of the intervention group at 15 out of 40.

Figure 1 summarises the students' responses to the initial questionnaire which they were given regarding their attitudes to weekly tests. Generally, it can be seen that the students in both groups had negative feelings about the weekly tests which they usually received in our school, with only one student in the control group and two students in the intervention group feeling relaxed before a weekly test.

This was backed up by comments made in the qualitative questionnaire section, in which several of the students stated that they did not like weekly tests. For example:

> I don't like tests because I'm not relaxed or happy. (Control group)

> Tests is too stressful for me and tiring for me so I don't like exams. (Control group)

> Test doesn't help me learn. (Intervention group)

Despite these negative attitudes to testing, students in both groups could see the advantages of weekly tests. For example:

> All tests have new vocabulary. It's very good for me. (Control group)

> I'm keen on tests when I'm studying at home. (Intervention group)

Figure 1: Responses to the question 'How do you feel before weekly tests?'

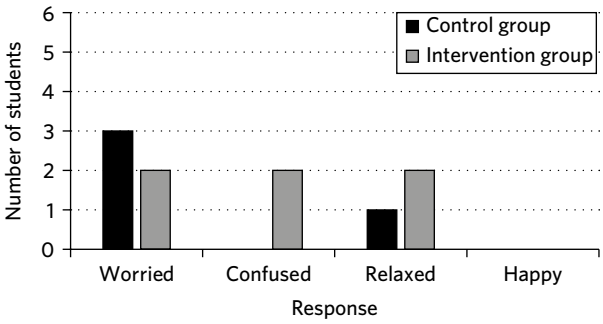
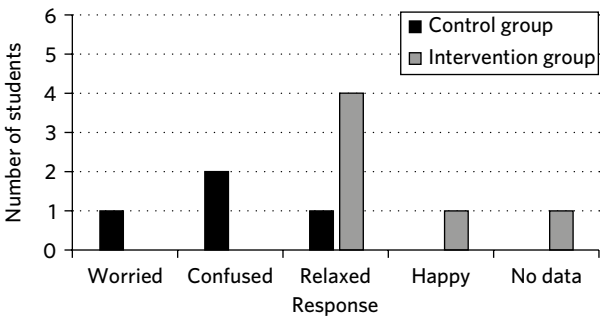


Figure 2 summarises the students' responses to the initial questionnaire about how they felt after weekly tests. To summarise, the activities completed on the first day revealed that, in general, the students in both groups felt worried or confused before weekly tests. These negative feelings were also felt by both groups after a weekly test, but the intervention group appeared to have slightly more positive feelings after a test. Both groups had a similar level of writing skills and the writing skills of the control group were slightly stronger than those of the intervention group. In terms of speaking, the intervention group as a whole had slightly stronger speaking skills than the control group.

Figure 2: Responses to the question 'How do you feel after weekly tests?'



Cycle 1: Intervention – The first six weeks

The programme of study then commenced. Although both classes followed a similar syllabus, different textbooks were used for each class. At the end of each week, the students were given a test. The intervention group was either given a speaking test or a writing test based on the grammar and vocabulary which the students had studied in that week. The control group was given a traditional discrete item test based on the work they had done that week. In Week 3, for example, both classes looked at past tenses. To evaluate the control group's understanding of this, they completed a discrete item test, which consisted of putting verbs in brackets into the correct form, either the past simple or continuous, whereas the intervention group had to produce a piece of writing describing their childhood. During this time, we also monitored how the students performed in speaking tasks during class time and copies were kept of the writing which they produced to see if there was any difference between their journal work, which is written work the students complete as homework on a weekly basis, and the written work which they produced in the mid-term test.

The end of Cycle 1

In order to gauge to what extent the students' productive skills had improved during the first cycle the students completed the following activities in Week 6:

1. A writing test, which comprised two tasks which were on similar topics (a description of the students' family and a description of a place) to the initial writing test. The writing results were compared against our baseline data to assess how their writing had improved.
2. A speaking test which consisted of two parts: an activity which required pairs of students to discuss the similarities and differences between some free-time activities based on visual cues such as a person cooking, a person reading, a person playing football, a person rock climbing etc., then to give an opinion on which would be better for a younger and older person. The second part was a long turn speaking task which required them to describe a person who they knew.

The students also:

1. Completed a questionnaire to see if their attitudes to language testing had changed.
2. Participated in a focus group session to see whether or not they felt the weekly tests had helped them in the mid-term test.

Cycle 1: Findings

This section will start by looking at the effect which the productive focus had on the intervention group by providing a summary of April's field notes. This will be followed by some observations made by both teachers during the mid-term test. We will then make a quantitative comparison of the students' writing and speaking by comparing their results in the mid-term test with the baseline data gathered at the start of Cycle 1. This will be followed by findings from the questionnaire and a summary of the key points stated in the focus group. This section will then conclude with two case studies.

Summary of April's field notes about the intervention group from Cycle 1

Week 1

The students didn't really know what to expect and asked a lot of questions during the test, e.g. 'but what do I write teacher?', 'how much do I write?'. When I responded that they should write as much as they could and to write whatever they felt best answered the questions I was met with bemused looks.

When the tests were marked and feedback given the students were quite upset, especially those who were used to receiving high test scores. Predictably none of the students performed well due to their basic, poorly structured and short answers. The students were reassured that this was just the first step on the road to improving their writing.

Week 2

Students were a little apprehensive before the test, so I adjusted my language and henceforth it was always referred to as a weekly 'review' rather than a 'test'. Some students showed a slight improvement, but one student in particular became quite angered by the process. He felt that the marking was arbitrary and did not agree with his result. I therefore had to spend quite a long time giving individual feedback on the errors and how they could improve. This was quite time-consuming but I felt it necessary to maintain students' buy-in.

Week 3

I decided to provide students with a copy of the marking guide, so that they could see what type of answer would receive the highest grades. This was very well received. Some students started to show further progress and enjoyed being ambitious with the new language. This was especially true of the speaking component – they started to understand how to maintain a conversation by asking follow-up questions. The stronger students also began to extend their spoken answers.

Weeks 4 and 5

By this stage, students felt more comfortable with the testing style and could see the benefits for themselves. They had expressed their thanks for this focus and had even told students from other classes how happy they were with their testing policy: 'Teacher you know other classes don't do writing practice every week, I told the other class that teacher April helps us to write.' Students this week held a conversation in English for 10 minutes, and they appeared engaged and animated throughout.

Week 6

This was the mid-term test week. The students felt relaxed and prepared before the test. All students looked relaxed during the writing component.

Teacher observations made during the mid-term test

It was extremely interesting to note the differences in atmosphere between the two classes. The students in the intervention class looked far more relaxed both before and during the mid-term test. The students in the control group vocalised their worries about the mid-term test on several occasions prior to the mid-term test, and during the test it was evident that they had problems managing their time. This was evidenced by the fact that several of the students were unable to complete the writing tasks in time. In the speaking tests, several of the students were visibly nervous and reported this to the teacher after the test.

Cycle 1 findings: Comparative data analysis

Writing

At the end of Cycle 1, there were improvements in the writing test scores of both groups in the mid-term test, compared to the students' performance in the initial writing test. Table 2 shows the average mark attained by each class in both tests.

Table 2: The average writing test results of both classes at the start and end of Cycle 1

Group	Initial writing test average mark out of 40	Mid-term writing test average mark out of 40
Control group	18	23.5
Intervention group	17.2	27.4

Whilst both groups achieved roughly the same average mark in the initial writing test and both groups showed improvement, it is noticeable how the intervention group performed better than the control group in the mid-term writing test and the average improvement of the intervention group was just over a third that of the control group.

Writing progress by criteria

Both groups showed improvement in all of the criteria as exemplified in Figures 3 and 4.

Speaking

There were improvements in the speaking abilities of both groups in the mid-term test, compared to the students' performance in the initial speaking test. Table 3 shows the average mark attained by each class in both tests.

Figure 3: The progress of the intervention group in writing by criteria

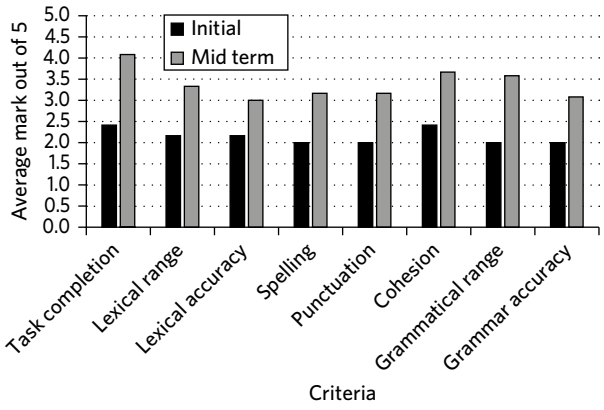


Figure 4: The progress of the control group in writing by criteria

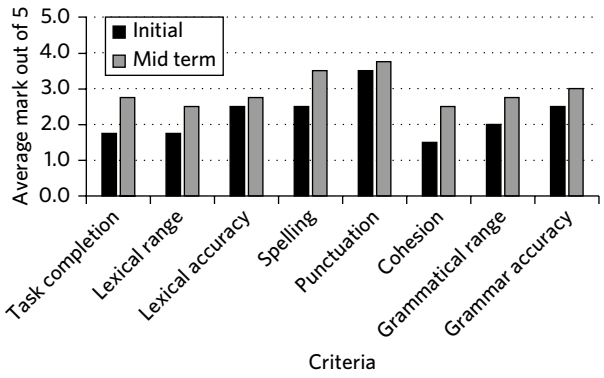


Table 3: Average speaking test results of both groups at the start and end of Cycle 1

Group	Average initial speaking test result out of 40	Mid-term speaking test result out of 40
Control group	12.4	19.8
Intervention group	13.8	25.8

Cycle 1 findings: Attitudes to testing

Questionnaire results

At the end of Cycle 1, we asked the students to complete a questionnaire to ascertain how they felt about testing and to see if their attitudes towards testing had changed at all over the course of Cycle 1. We asked them how they felt before and after weekly tests. The results were compared against the results of the initial questionnaire given at the start of Cycle 1 and are summarised in Figures 5 and 6.

At the start of Cycle 1, three of the four students in the control group had negative feelings both before and after a test and this had not changed by the end of Cycle 1. However, in the intervention group it seemed that there had been a shift in opinion. Before Cycle 1 commenced two of the students had said that they felt worried and two of the students had said that they felt confused before the

weekly tests, but by the end of Cycle 1 only one student felt confused and one student felt worried. The other four students of the intervention group felt either relaxed or happy, which suggested that they had more positive feelings towards weekly testing than they did at the start of Cycle 1. At the start of Cycle 1, the majority of the students in the intervention group reported feeling relaxed after a weekly test and by the end of Cycle 1 this had shifted to feeling happy.

Figure 5: Comparison of how both groups felt before weekly tests before and after Cycle 1

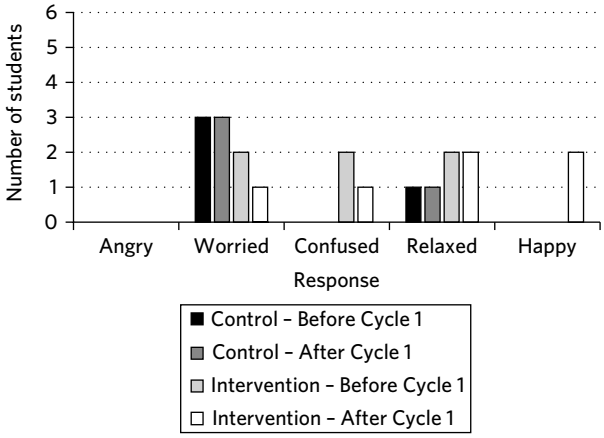
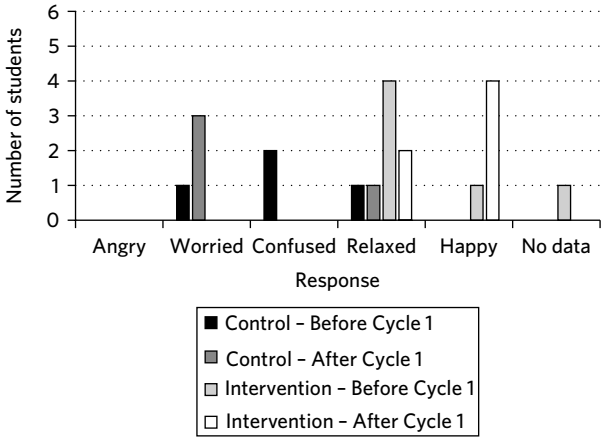


Figure 6: Comparison of how both groups felt after weekly tests before and after Cycle 1



We also asked the students to complete a questionnaire to ascertain how they felt both before and after the mid-term test. The results are summarised in Figures 7 and 8. The results of this question revealed that all of the control students felt worried about the mid-term test, whilst only one of the students in the intervention felt worried about it. Two of the students in the intervention group felt confused, but the remainder felt relaxed.

It was interesting to note that only one of the students in the control group had positive feelings after the mid-term test; the remaining three students continued to feel worried. In the intervention group, however, the students who answered this question reported having more positive feelings and feeling either happy or relaxed.

Figure 7: Results of the question 'How did you feel before the mid-term test?'

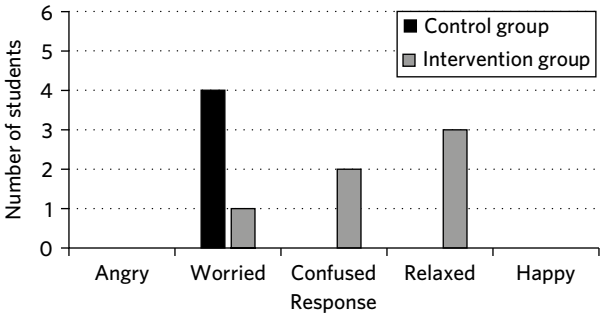
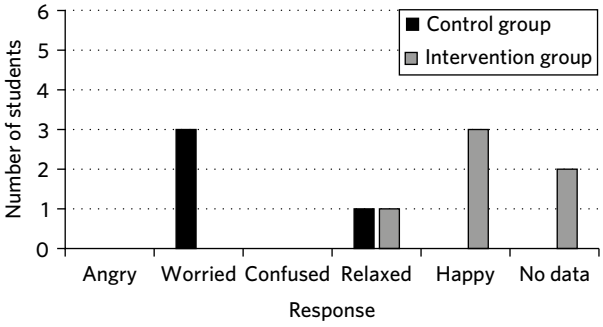


Figure 8: Results of the question 'How did you feel after the mid-term test?'



The focus group

In order to better understand the students' attitudes to tests, we also carried out a focus group session with each of the groups. The focus group for the intervention group was carried out by Ceri and all six students from this group participated, with Ceri taking the focus group for the control group in which all four students participated. Both focus groups lasted for approximately 20 minutes and took place during class time. The comments made by the students broadly mirrored the views they had expressed in the questionnaires.

The students in the intervention group said that they enjoyed doing weekly tests and that these had helped the students to prepare for the mid-term test. The extracts below are representative of the group as a whole.

> I like the weekly test because it help us in the mid-term. (Ali)
>
> I like the writing and spelling because I'm not good at it and it help me improve. (Hamad)

Interestingly one of the students had noted in his questionnaire at the start of Cycle 1:

> I like the maltbal choose [multiple choice] question. (Abdulaziz)

However, in the focus group he stated:

> I like writing because it help me to improve my life and use my English.

Another student had reported in the initial questionnaire at the start of Cycle 1 that 'I'm not keen on about writing' but in the focus group he said:

> Yes I feel I have improved my writing, I can write more and more now. (Abdulatif)

By contrast, the students in the control group were less positive about tests. Whilst they could see the benefits of tests by making comments such as:

> I like to know my score. It's important. (Yoon)

several of them reported negative feelings such as:

> I don't like tests. I feel stressed and nervous . . . I feel happy when the test finish. (Cristian)

Overall, the results of the questionnaires and the focus group showed that the students in the intervention group had more positive feelings towards tests and felt that the tests had brought about greater benefits to them than the students in the control group.

Cycle 1 findings: Case studies

Writing case study

Yoon was a student from the control group and his performance in the mid-term writing test compared to his journal work was representative of the control group as a whole. He was a South Korean student in the control group and achieved 24 out of 40 in the mid-term writing test. In the journal work he produced during Cycle 1, he showed a greater ability to use language which we had studied in class and to use adjectives to add interest to his writing. He showed quite a good range of vocabulary using lexis such as 'surrounded by' and 'inland areas'. In his journal work he would typically try to give detailed pieces of writing and expand his ideas, making use of longer, more complex sentences by linking ideas together. Typically in his journal work he was capable of writing around 220 words. In the mid-term test however, his writing was much shorter. He also used much shorter sentences, using linkers to start sentences rather than joining them together. He also used a much narrower range of vocabulary compared to the journal and produced writing of only around 130 words.

Speaking case study

Miyuki was a Japanese student who was a member of the control group and her reactions to the speaking test were representative of the reaction of the control group as a whole. Miyuki started the course with relatively low speaking skills and achieved the lowest score of the whole class in the initial speaking test. She was unable to formulate ideas and spent most of the initial speaking test pausing to search for the language she needed. One of her priorities was to improve her speaking skills and as her confidence grew, she showed good progress in this area. She responded well to Dogme[1] lessons, which are 'conversation-driven' (Meddings and Thornbury 2009:8) and based around natural conversations. She would often ask 'Speaking today teacher?' with a big smile on her face.

In the mid-term speaking test, however, she was visibly nervous before the test and looked very unhappy during the test. Once again, she found it difficult to speak under the pressure as she had not been prepared for it, and after the test she said that she didn't enjoy the speaking test because it was too stressful.

Overall it seemed from Cycle 1 that the introduction of productive skills testing had a positive effect not only on the intervention group's test results but also on their attitudes to testing. Whilst there was also an improvement in the writing and speaking skills of the control group, this was not as significant as the improvement in the intervention group, and in contrast to the intervention group, the control group continued to have negative feelings towards testing.

Cycle 2

In order to seek confirmation of the findings of Cycle 1 and to implement some changes which we felt necessary, we carried out a second cycle of research which also lasted for six weeks.

Changes we decided to make

In Cycle 2, we decided to keep the fundamentals of the project the same. We did, however, decide to make certain changes to the process. First of all, we decided to ask the students of both classes the following question at the end of Cycle 2: How much do you feel that the weekly tests helped you to prepare for the writing part of the end-of-term test? Students were able to respond to this question on a scale from 1 to 5, with 5 signalling that they felt the weekly tests had helped them fully and 1 signalling that the weekly tests had not really helped them.

Secondly, the students in the intervention group had complained during Cycle 1 that they did not understand how their weekly test scores were calculated. As the assessment criteria had been written for the teacher to use, these were complex for the students to understand. We therefore decided to create a simplified version which the students would be better able to understand.

Finally, we decided to limit the focus of our study to writing. This was for three main reasons. Firstly, the process of giving speaking tasks to the intervention group was quite time-consuming and in general speaking was one of their strengths. Secondly, as the intervention group still had weaknesses in writing, we felt it would be better to focus in on this area. Thirdly, as the control group required some additional help in improving their speaking skills, which was carried out during class time, we felt that this would create too much of a variable in our results as more class time was given to speaking during the second cycle for the control group compared to the intervention group.

[1] Dogme is a communicative approach to language teaching which focuses on conversational communication between the teacher and learners.

The students

In the second cycle, there were some changes to the classes. The majority of students in the intervention group were Muslim and as the end of Cycle 2 clashed with Ramadan, this meant that some of the students decided to return to Saudi Arabia. This resulted in the intervention group being much smaller, consisting of two students from the original group and another student who joined the class who was from Greece. In the control group, the original four students remained and they were also joined by a student from Brazil. The teachers of the two groups stayed the same.

The first day

At the start of the cycle, the students completed an initial writing task which took the form of a short academic essay which asked the students to discuss whether or not it was better to study English in an English-speaking country or in their home country. The new students also completed questionnaires. Once again, this yielded some baseline data, which is represented in Table 4.

Table 4: Baseline writing results at the start of Cycle 2

Group	Average baseline writing score at the start of Cycle 2 out of 40
Control group	20.8
Intervention group	24.2

Table 4 represents the average mark achieved by both groups in the initial writing test given at the start of Cycle 2. In contrast to the baseline writing marks at the start of Cycle 1, there was a greater variance in writing skills between the two groups. This was probably caused by the effect which Cycle 1 had had on their writing skills. For the control group the lowest mark in the initial writing was 15.5 out of 40 and the highest mark achieved was 31.5 out of 40, which was achieved by the Italian student. The range of marks achieved by the intervention group was slightly narrower with the lowest mark at 16 out of 40 and the highest mark at 31 out of 40.

Cycle 2: Intervention – the second six weeks

The programme of study for Cycle 2 then commenced. As in Cycle 1, both classes followed a similar syllabus, although different textbooks were used for each class. At the end of each week, the students were given a test. The intervention group was given a writing test, which the teacher wrote, based on the grammar and vocabulary which the students had studied in that week. The control group was given a traditional discrete item test based on the work they had done that week. As in Cycle 1, copies were kept of the writing which they produced to see if there was any difference between their journal work and the written work which they produced in the end-of-term test.

In short, at the start of Cycle 2 we decided to focus our study on writing and we redesigned the marking scheme to make it more student friendly. Cycle 2 was then carried out in the same way as Cycle 1 with an initial test, a productive writing test each week for the intervention group, a discrete test for the control group and finally both groups completed the same writing test at the end of the cycle. Once Cycle 2 had finished, the students completed a questionnaire and participated in a focus group.

Cycle 2: Findings

In this section we will start by looking at the effect which the productive focus had on the intervention group by providing a summary of April's field notes. We will then make a quantitative comparison of the students' writing results compared to their baseline results. We will then look at the results of the questionnaires and the focus group. This section will conclude with a case study of one of the students in the intervention group.

Summary of April's field notes from Cycle 2

Week 7

Some students started to feel so confident in their own progress that they sat an external *IELTS* examination.

Weeks 8–11

As the students were now fully aware of the testing strategy I no longer had to provide lengthy written feedback and could elicit from them areas of improvement. As I pushed them to improve further, inadvertently my marking did become stricter.

Week 12

Students felt prepared and confident before the end-of-term test, especially with regard to writing and speaking components. The core group of students achieved good results and asked if they could follow the same procedure in the forthcoming term as they felt it had greatly benefited them.

Cycle 2: Data analysis

Writing

At the end of Cycle 2, there were improvements in the standard of writing of both groups in the end-of-term test, compared to the students' performance in the writing they had done at the start of Cycle 2. Table 5 shows the average mark attained by each class in both tests.

Table 5: Average writing test results for both groups at the start and end of Cycle 2

Group	Average mark out of 40 in the initial speaking test	Average mark out of 40 in the mid-term speaking test
Control group	20.8	24.1
Intervention group	24.2	29.5

At the start of Cycle 2, the intervention group had stronger writing skills than the students in the control group, with the former group achieving an average of 24.2 out of 40 and the latter group achieving 20.8 out of 40. Both groups showed an improvement in their writing skills with the average mark increasing from 20.8 to 24.1 in the control group, an increase of 3.3 points, and the average mark increasing from 24.2 to 29.5 in the intervention group, an increase of 5.3 points. These differences are illustrated in Figures 9–11.

Figure 9: Average improvement of both groups by the end of Cycle 2

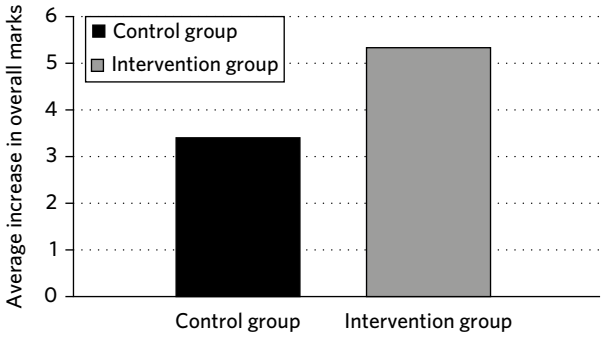


Figure 10: The progress of the intervention group in writing during Cycle 2 by criteria

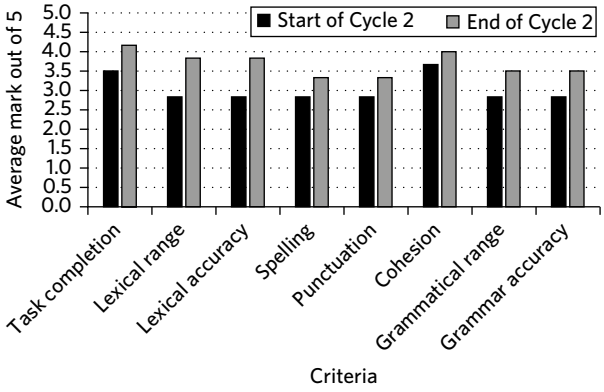
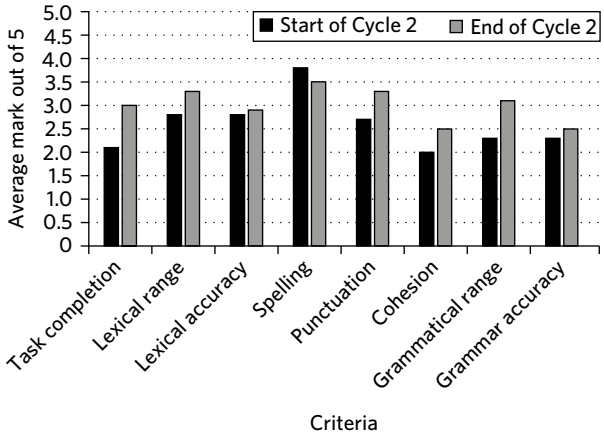


Figure 11: The progress of the control group in writing during Cycle 2 by criteria



The intervention group made progress in all areas, although the progress was less marked than in Cycle 1. As a whole the group made its biggest improvements in lexical range and accuracy, followed by grammatical range and accuracy, and task completion. In contrast to Cycle 1, not only did the intervention group achieve a marked improvement in grammatical and lexical range but also in accuracy. The smallest improvement for the intervention group as a whole was in cohesion. In the control group the biggest improvement was in the area of task achievement. In contrast to Cycle 1, the control group also saw an improvement in grammatical range, although the improvement in lexis and accuracy was much smaller. The control group performed worse in the area of spelling and this is perhaps because the students were unused to doing writing tasks and during the test many of them managed their time badly, which resulted in them not having sufficient time to check their work for spelling errors.

Cycle 2 findings: Attitudes to testing

Questionnaire results

At the end of Cycle 2, students were asked to complete a questionnaire about their attitudes towards testing. It is clear from Figure 12 that there was only a small difference to the feelings of the control group about weekly testing, whereas the intervention group moved towards more positive feelings.

Both Figures 13 and 14 show that the control group was more worried about the end-of-term exam both before and after the exam had taken place, whereas the intervention group on the whole felt more happy and relaxed. The intervention group also felt that the productive weekly tests had better prepared them for the writing task in the end-of-term exam, as exemplified by Figure 15.

Figure 12: Comparison of how the students felt about weekly tests before the action research commenced and at the end of Cycle 2

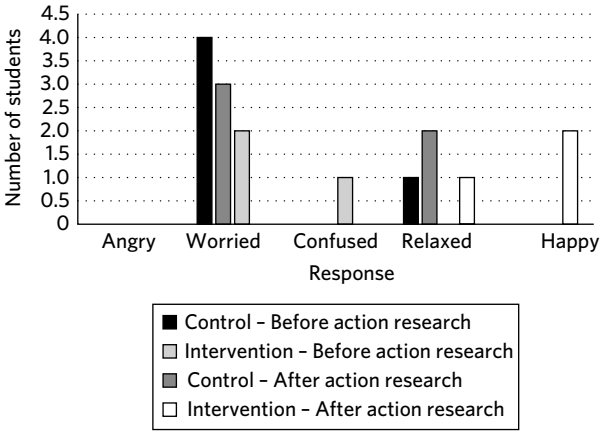


Figure 13: Responses to the question 'How did you feel before the end-of-term exam?'

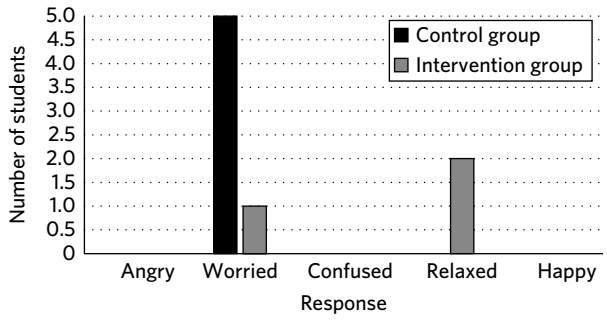


Figure 14: Responses to the question 'How did you feel after the end-of-term exam?'

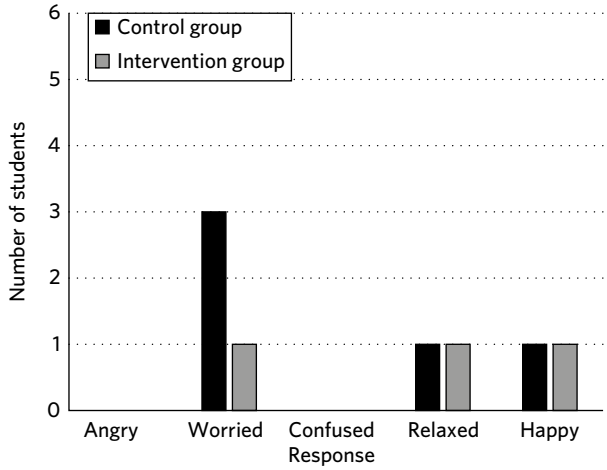
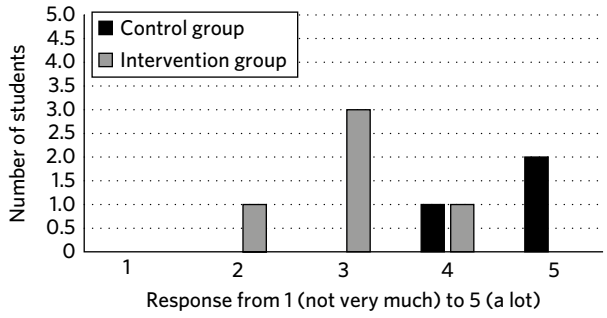


Figure 15: Responses to the question 'How much do you think that the weekly tests helped you to prepare for the writing you had to do in the end-of-term exam?'



Focus group

Finally, we carried out a focus group to gain more insight into what the students felt about tests. The focus group for the intervention class was carried out by Ceri and consisted of the three students who had taken part in Cycle 2, and the focus group for the control class was carried out by April and consisted of the five students who had taken part in Cycle 2. Both focus group sessions lasted for around 15 minutes. The samples given below are representative of the feelings of the two groups (the first two quotes are from the control group and the subsequent quotes are from the intervention group).

I like to know my score. It is important. (Yoon)

I don't look at my test again . . . Just the book. (Diana)

This will be a big help to IELTS . . . I feel I am ready now for IELTS. (Abdullatif)

We don't say tests, we say reviews. Yes they are very important to help us to know what level we are and how to be better. Also teacher April give us advice how to improve. (Abdulaziz)

I keep all my writing together and look at it many time . . . I can see I am stronger in my English. (Walid)

Cycle 2 findings: Intervention group case study – Walid

Walid came to the school at elementary level, needing to achieve 4.5 in *IELTS*. He had a relatively short timescale in which to achieve this considering that students require on average three months of study in order to improve by 0.5 of a band. His grammar knowledge and lexical range were at A2 level of the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001) at the start of the project. However, his spelling was very weak and he had never written extensively in English beforehand. Initially, Walid was excited about the project as he felt it could help him to achieve his goal. However, he quickly became frustrated with the productive style of testing as he found this more difficult than traditional testing. He questioned the grading of his written work and voiced the opinion that he felt the marking was done arbitrarily.

Taking this feedback on board, we created some student-friendly marking criteria, which showed what level of detail was needed for them to achieve each band score. Walid responded really well and made a concerted effort to meet the criteria in his written and spoken English.

Instead of dreading writing extensively in English, he began to relish the opportunity to show off what he had learned that week. He even began writing extra pieces at home. His writing skills improved dramatically, so much so that he achieved 4 in *IELTS* Writing within nine weeks.

He really enjoyed the focus on producing and using the language. His confidence grew and he began socialising in English outside of the classroom. He stated that the style of the weekly tests helped him to remain calm when under external exam conditions. He achieved overall *IELTS* 4.5 (4 Writing and 5 Speaking) nine weeks into the project.

Discussion of key findings

Having explained how we conducted our action research and outlined the key findings after each cycle, we will now discuss some of the key findings of the project.

The central aim of this project was to discover whether or not a greater focus on productive skills in weekly tests would have any effect on our students' productive skills in summative assessment. Whilst it is difficult to draw any concrete conclusions about this as there are many factors

which affect a student's ability to do well in a productive skills test, we do feel that the greater focus which was given to speaking and writing tasks in the weekly tests of the intervention group had a positive impact on their ability to perform well in both the mid-term and end-of-term tests.

In terms of writing, both the control group and the intervention group started the course with more or less the same level of writing ability. Although both groups showed an improvement in their writing by the mid-term test, the intervention group outperformed the control group and as a class showed an average improvement in writing of 9.6 points compared to the average improvement of the control group's 6 points. Whilst there were improvements in all areas of both groups' writing, there was a much greater improvement in the intervention group's performance in the areas of task completion, grammatical range, cohesion and lexical range. The improvement in accuracy was not as great. One of the reasons may have been that as the students were used to performing writing tasks in class on a regular basis, their confidence had grown and they were therefore more willing to take chances, experiment with language and to push themselves. By the end-of-term exam, the experimental group was able to produce on average 220 words, but the control group produced work of on average 160 words. This growth in confidence, however, was not always an easy one, especially at the start of the course. This is revealed in April's field notes, where she mentions the problems which she encountered at the start of Cycle 1 in terms of the students feeling unsure about what they were expected to do and the frustration the students felt when they believed they had underachieved and did not understand how their tests had been marked.

In contrast to the first cycle, there was a much greater difference between the average scores of each class when they completed the first piece of writing which provided our baseline data for Cycle 2. An underlying reason was probably that the students in the intervention group were used to producing writing under test circumstances and had greater confidence in themselves. As in Cycle 1, there was an increase in the test scores of both groups. In Cycle 2, however, the difference in improvement between the two groups was not as great as it had been in Cycle 1. Nevertheless, the improvement of the intervention group was slightly higher than that of the control group and in Cycle 2, the intervention group made improvements in accuracy. It was also interesting that the control group witnessed a decline in spelling. This seems to have been caused by the inability of the control group to manage their time very well, which meant that several of the students had to rush their writing and did not have enough time to check their writing for spelling errors.

Due to time constraints, the decision was taken at the end of Cycle 1 to limit the scope of our study to just writing. Nevertheless, the results of the speaking tests carried out in Cycle 1 mirrored the improvements which were seen in writing. At the start of Cycle 1, the intervention group showed that they had slightly higher speaking skills than the control group, by just over one point. At the end of Cycle 1, both groups had once again improved but the gap between the two groups had widened with the intervention group scoring an average of 6 points more than the control group. As with the writing, the intervention group showed a greater willingness to make use of grammatical structures and lexis which had been studied during the course.

Another factor which we feel was important is that the intervention group had far more realistic expectations about what they could expect to achieve in both the mid-term and end-of-term test. As the control group received discrete item tests on a weekly basis, they typically received a high mark each week, but their marks in the mid-term and end-of-term tests dropped considerably. The results of the intervention group were far more consistent over the course of the term.

As we mentioned in our literature review, summative assessment can be a stressful experience for students (Harris and McCann 1994). We feel that our observations of the experimental group and student feedback suggest that allowing students the opportunity to practise writing in tests on a weekly basis can help them become more accustomed to the pressures of writing in summative assessment and goes some way to relieving the stress of summative assessment.

One final area which we explored was the students' attitudes to tests. It was extremely interesting to note that before the action research, the students in the intervention group had mixed feelings about tests, but by the end of the project some of the students in this group had a more favourable attitude towards tests. On the other hand, the control group started the project with generally negative attitudes towards tests and their attitudes towards testing remained broadly unchanged. As we have already mentioned, Lambert and Lines (2000) make the point that oral assessment can be motivating and we certainly found this to be the case with the intervention group in Cycle 1.

Previously in this report we mentioned that Whitehead and Manassian (2013) believe that testing should go beyond passive understanding and focus less on what students know about language and more on the student's ability to use the language. Rea-Dickins also states that 'assessment procedures which only yield scores or grades do not adequately fulfil the needs of classroom based assessment' (2000:385). We found that the greater focus on productive skills helped April gain a deeper understanding of the areas which the students in the intervention group needed to work on. We also mentioned that tests can have an effect on teaching and, due to greater insight which April got into her students' productive skills through greater use of productive tests, she was able to better teach to the needs of her students, an example of positive washback. Furthermore the individual students in the intervention group developed an understanding of the areas which they needed to work on to improve their writing and speaking skills.

Having carried out our project, we feel that weekly tests should comprise a mixture of different task types and

include both discrete items and productive tasks, with greater emphasis given to the latter. Due to the nature of action research, certain factors were beyond our control such as the small number of students in each class and the variations in nationality mix between the two classes.

The next steps

One of the benefits of action research is that it is an ongoing process and we intend to continue with this project into the future in order to gain a more profound understanding and further validate our findings. We intend to repeat this project with two more groups, this time at intermediate level. This time we will adopt a different strategy in each cycle. In the first cycle, the students of both groups will complete discrete item tasks in their weekly test. In the second cycle the students of both groups will complete weekly tests with a more productive focus. This will allow us to compare the effect which discrete item testing and productive skills testing have on the same students. As has been mentioned one of the drawbacks of testing productive skills is that assessment is conducted by a person who has a subjective understanding; in other words there can be variances between the way teachers mark writing and speaking tests. In order to remove this possible variance, we will implement a system of second marking.

Reflections

Whilst participating in an action research project was time-consuming and challenging, we feel that it had benefits for both of us and was an enjoyable experience. On the downside it was time-consuming. Most coursebooks come with tests which can easily be used as weekly tests and as we mentioned earlier discrete item tests are easy to mark. During the course of the experiment Ceri was easily able to mark the tests in his class and preparation time was minimal. On the other hand, April had to spend approximately 30 minutes each week creating a test and also had to spend approximately 20 minutes per student each week marking tests and writing feedback, which was a considerable workload.

It was also challenging to fulfil three different roles: being the teacher, the assessor and researcher. At times it was also difficult to remain objective, both in terms of assessing our students and carrying out the research. We felt, however, that the marking scale we had created in the planning stage and the decisions we made about how we would evaluate the effectiveness of our research at the planning stage helped us to maintain a greater degree of objectivity than we would have had we not considered these issues initially.

Despite these challenges, we both felt that the project had many positive aspects. We feel that by seeking the views of the students and involving the students in the project we both became closer to our students and developed a greater understanding of them in a way which may not have been possible without the project. As we have mentioned, not only did April develop a greater understanding of her students' needs, her students also became more involved in understanding the areas which they needed to work on.

We also felt that we gained a lot from our background reading and through experimenting with tests. We feel that we have learned a great deal about assessment and become more knowledgeable in this area.

Finally, in the time since Cycle 2 finished, we have both started teaching different classes and both of us have been placing greater emphasis on productive tasks in the weekly tests of these classes. We have continued to make field notes and have noticed that the productive element continues to engage and motivate students. In terms of our institution, our colleagues have shown interest in our action research project and we have delivered a training session to them on adding a productive focus to weekly tests. Some of our colleagues have also adapted the marking scale which we designed for our project to make it suitable for their own classes.

Overall, this has been a positive, yet challenging, experience for the both of us and we would thoroughly recommend action research as a systematic way of investigating and improving classroom practice.

References

Anderson, G (1998) *Fundamentals of Educational Research*, London: Sage

Baxter, A (1997) *Evaluating Your Students*, London: Richmond.

Borg, S (2010) Doing good quality research, *JACET Journal* 50, 9–13.

Burns, A (2010) *Doing Action Research in English Language Teaching*, Abingdon: Routledge.

Campbell, A McNamara, O and Gilroy, P (2004) *Practitioner Research and Professional Development in Education*, London: Paul Chapman.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.

Douglas-Brown, H (2004) *Language Assessment Principles and Classroom Practices*, London: Pearson Longman.

Gong, B (2010) *Considerations of Conducting Spoken English Tests for Advanced College Students*, available online: www.iaea.info/documents/paper_4d34afa.pdf

Harris, M and McCann, P (1994) *Assessment*, Oxford: Macmillan.

Heaton, J B (1988) *Writing English Language Tests*, London: Longman.

Hendricks, C (2006) *Improving Schools through Action Research*, Boston: Pearson.

Hopkins, D (2008) *A Teacher's Guide to Classroom Research*, Buckingham: Open University Press.

Hughes, A (2003) *Testing for Language Teachers*, Cambridge: Cambridge University Press.

Lambert, D and Lines, D (2000) *Understanding Assessment: Purposes, Perceptions, Practice*, London: Routledge.

Meddings, L and Thornbury, S (2009) *Teaching Unplugged*, Surrey: DELTA Publishing.

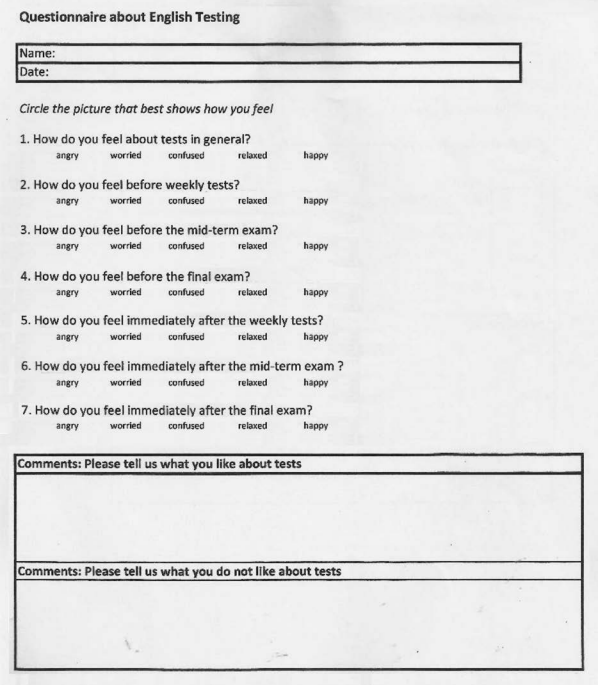
Rea-Dickins, P (2000) Classroom assessment, in Hedge, T (Ed) *Teaching and Learning in the Language Classroom*, Oxford: Oxford University Press, 375–400.

Salvisberg, J (2014) Formative assessment of oral skills: Does the end justify the means? in Pattison, T (Ed) *IATEFL 2013 Liverpool Conference Selections*, Kent: IATEFL, 182–184.

Tisdall, E K M, Davis, J M and Gallagher, M (2009) *Researching with Children and Young People: Research Methods, Design and Analysis*, London: Sage.

Whitehead, R and Manassian, S (2014) How to write a good test, in Pattison, T (Ed) *IATEFL 2013 Liverpool Conference Selections*, Kent: IATEFL, 186–187.

Appendix 1: The initial questionnaire



Questionnaire about English Testing

Name:
Date:

Circle the picture that best shows how you feel

1. How do you feel about tests in general?
angry worried confused relaxed happy

2. How do you feel before weekly tests?
angry worried confused relaxed happy

3. How do you feel before the mid-term exam?
angry worried confused relaxed happy

4. How do you feel before the final exam?
angry worried confused relaxed happy

5. How do you feel immediately after the weekly tests?
angry worried confused relaxed happy

6. How do you feel immediately after the mid-term exam ?
angry worried confused relaxed happy

7. How do you feel immediately after the final exam?
angry worried confused relaxed happy

Comments: Please tell us what you like about tests

Comments: Please tell us what you do not like about tests

Appendix 2: The initial writing test

Imagine that you are writing an email to a friend and you want to tell him or her about a place you have recently visited. Make sure that you include a description of the place you visited, what you did while you were there, who you went with and why you enjoyed or didn't enjoy visiting this place.

Try to write between 75 and 100 words.

_____
_____
_____
_____
_____

Write a short paragraph about your family. Try to include as much information as possible about them and write between 75 and 100 words.

_____
_____
_____
_____
_____

Appendix 3: The questions used in part 1 of the initial speaking test

What's your name?
How old are you?
Where are you from?
Tell me about your family? Have you got any brothers or sisters?
Do you have a large family?
What was your favourite subject at school? Who was your favourite teacher? Why?
What subject didn't you like? Why?
What do you enjoy doing in your free time? Do you play any sports? Do you play a musical instrument?
Have you travelled abroad before? What other countries have you visited?
Tell me about your last summer holiday? What did you do last weekend?
How does Cardiff compare with your home town/city?
If you had three wishes, what would you ask for?
If you could live anywhere in the world for a year, where would you go?

Investigating attitudes to traditional and online homework using action research

ANDREW TAYLOR ST GILES LONDON CENTRAL

Introduction

One of my key assumptions about my learners was that most of them, being digitally literate, would have a more favourable disposition towards doing homework online. An opportunity arose to test these assumptions. The language school where I work was keen to promote their website called E Learning as an online self-study learning tool. E Learning is a multilevel website. It provides a range of things to study, and in most cases provides automatic grading of exercises completed. Figure 1 shows the range of what can be studied in E Learning and Figure 2 shows the types of tasks that are generally available in each section.

I became interested in how my learners would respond to a change in homework practices, where they would have the opportunity to use both E Learning as well as completing traditional paper-based homework tasks. North and Pillay (2002:144) point out that as the use of the internet for English language resources increases, so does the need to re-examine homework practices. Would the use of online learning tools have an impact on my learners' homework practices?

Through the use of action research, my aim was to explore my learners' attitudes to using the E Learning platform and paper-based tasks for doing homework. The research was conducted over two 5-week cycles of mixed methods research. My class consisted of 12 adult learners who were mostly in their 20s and had a B1 Common European Framework of Reference for Languages (CEFR, Council of Europe 2001) level of proficiency (intermediate) in English as a foreign language. I had a broad mixture of nationalities including Japanese, Korean, Colombian, Italian and Turkish. The class was a continuous enrolment General English class, which meant that there could potentially be new students joining every week, as well as other students leaving the class.

Online versus paper-based homework

Doorn, Janssen and O'Brien (2010) discovered that 55% of learners surveyed about traditional versus online homework said they preferred online homework, which was not an overwhelmingly positive response to doing online

Figure 1: Range of E Learning topics

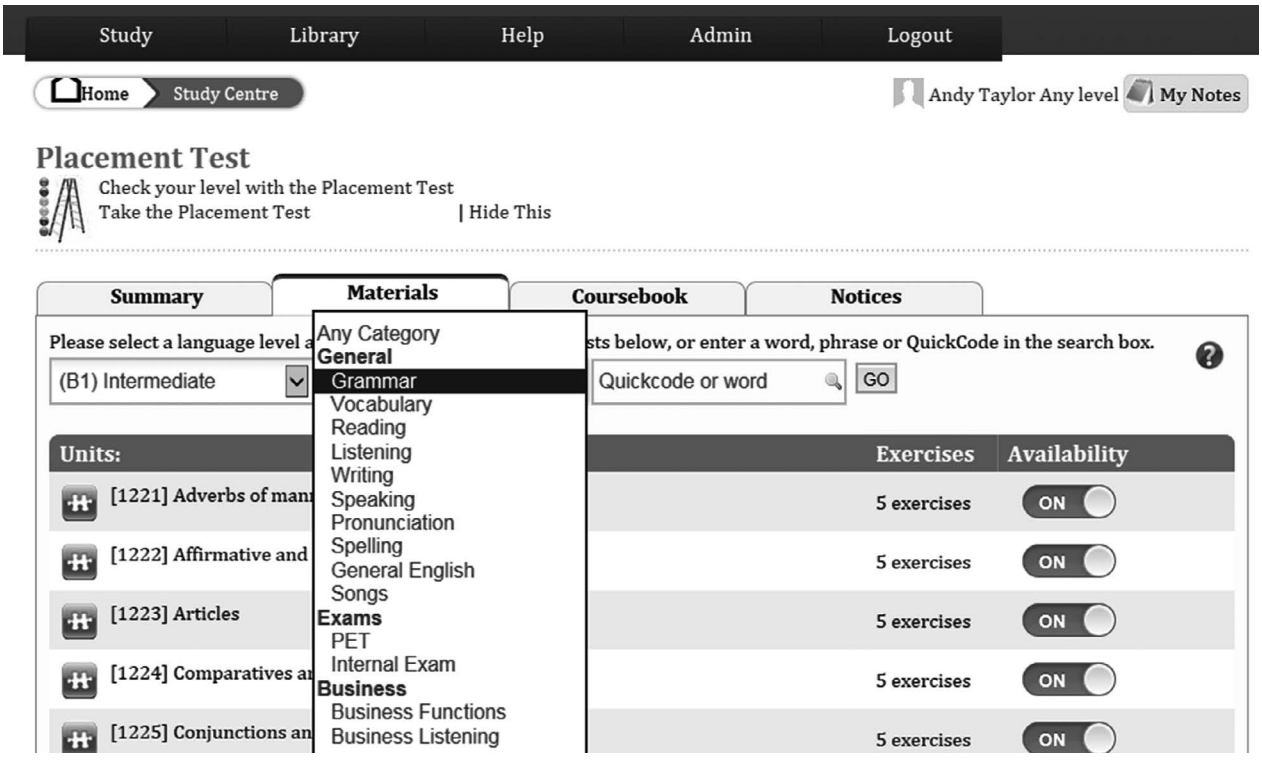
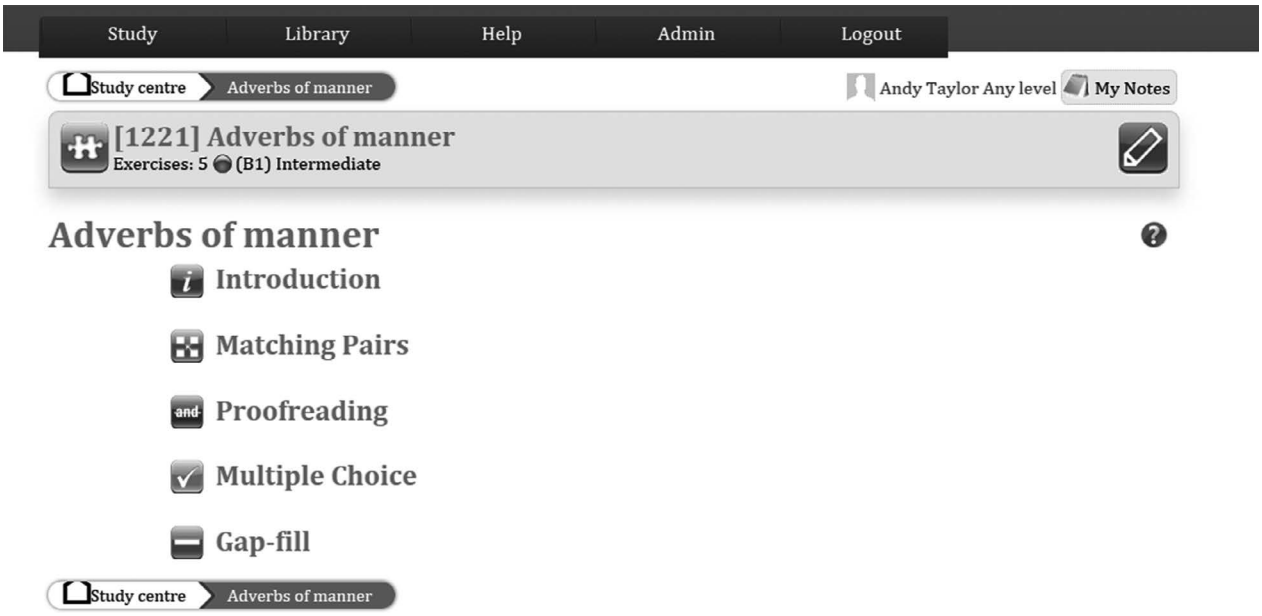


Figure 2: Range of E Learning task types



homework. Opinions were more divided when it came to learner perceptions of whether or not they learned more or less when studying online. It should be noted, however, that the study was carried out within a non-EFL context with university economics learners.

The homework medium does not appear to be important for motivated learners. Wooten and Dillard-Eggers (2013) show that motivated university learners would do homework regardless of whether it was online or paper based. Learners with less motivation would have to make more of an effort with online homework tasks. They surveyed classes that required online homework as part of the course and classes where it was optional. Learners who opted to do online homework did not have a more favourable perception of it than those who had to do compulsory online homework. The results did not reflect a strong endorsement of the benefits of technology. Again this was research conducted in a non-EFL university context, with accounting classes.

Wooten and Dillard-Eggers' findings were consistent with Doorn et al (2010), in that highly motivated learners were the most satisfied with online homework. In Wooten and Dillard-Eggers' research, they also say that both users and non-users of online homework preferred doing online homework tasks to using pencil and paper.

There were factors to consider for my research, such as the type of tasks that the learners were doing (regardless of the medium) and how they meet my learners' needs. North and Pillay (2002) suggest that a student's willingness to do homework might also be affected by the nature of the task; on whether it is challenging or engaging enough.

Both Painter (2003) and Hong, Wan and Peng (2011) talk about the importance of the relevance as perceived by the learners themselves of the homework tasks they are doing.

Painter also supports the idea that student homework should match the preferences of the learner. For instance, preferences for online versus paper-based tasks, or task types that lend themselves to online delivery. Could the E Learning platform reflect some of these preferences, especially if it reflected how my learners used technology outside the classroom?

The use of web-based technologies to complete homework tasks creates an overlap between two different notions of authenticity. On the one hand, the homework *task* itself might be seen as authentic if it reflects something learners do in everyday life. For example, using Facebook to make social arrangements or using an online template to prepare a CV for a part-time job that they could apply for in real life. The use of E Learning might make similar general claims to having task authenticity for writing exercises such as the CV example, but would score low on task authenticity for reading or listening exercises that are graded to specific levels and might have subject matter which is removed from the context of everyday student life. The other type of authenticity which the use of E Learning might have claims to is *functional authenticity*. Buendgens-Kosten (2013:280) defined this as learners doing an activity which mirrors the 'ordinary practices of the culture'. In this context, it is the culture of the student that is being referred to, not the target culture i.e. English language users living in London. Functional authenticity takes student activities and interests outside the classroom into consideration. E Learning could therefore claim to have task authenticity if it is something that they could *do* in their real lives (i.e. write a CV), but also functional authenticity if the web-based technology that they use to complete the task is something that they would use regularly outside the classroom in real life or student culture (i.e. writing a CV online).

Both Doorn et al (2010) and Wooten and Dillard-Eggers' (2013) studies were based on closed non-EFL classes in a university setting, where the composition of learners was stable. One of the reasons for my research was to see if I could carry out a similar study on my continuous enrolment General English class, as there seemed to be a lack of similar homework studies in the EFL sector.

The study

The purpose of the research was to investigate my learners' attitudes to both doing homework on the school's E Learning platform, and doing paper-based homework. My initial research question was: How does the use of an E Learning platform impact on learner attitudes to doing homework?

This question was loosely based on research issues which resonated with me from two other studies. Do learners find online homework more convenient or useful than traditional homework (Doorn et al 2010:2)? Are learner perceptions of doing online and doing more conventional homework different (Wooten and Dillard-Eggers 2013:190)? However, my initial research question evolved into a question with a slightly narrower focus for the second cycle. The second research question, and my reasons for having one, will be discussed in what I did for Cycle 2.

The research was conducted at St Giles London Central, which is the private language school where I work as a full-time EFL teacher. My action research was spread over two 5-week cycles. I chose this length of time for each cycle as the duration seemed long enough to engage learners in a number of online and paper-based tasks. Participants in the investigation included learners from my intermediate (or B1) General English class of 12 learners in a continuous enrolment environment.

Cycle 1

I taught my class for 3 hours a day every morning from Monday to Friday, or 15 hours a week. During a typical teaching week, an average of two E Learning assignments were given out, while other tasks were paper based. Out of the 25-day cycle, 10 homework tasks were E Learning assignments: two were a choice between E Learning and paper-based tasks, and the rest were paper based.

In order to gain deeper insights into the impact of E Learning on learner attitudes to doing homework, I adopted a mixed methods approach, collecting both quantitative and qualitative data.

During Week 1 and Week 3 of the cycle, my learners did an online survey (Appendices 1–2). I designed the questions to elicit student attitudes to homework in general, as well as attitudes to doing it with E Learning and with paper and pen. A Likert scale was used to capture most of their responses. Nine learners completed the first survey (out of nine) and 12 completed the second one (out of 12), with the same six learners completing both. Therefore I used the information generated by those six learners to study how attitudes may have altered over time.

In Week 5, on the final day of the first cycle, I conducted nine 4–5-minute semi-structured one-to-one interviews with each learner in the class. After learners gave oral consent, these interviews were recorded and then transcribed. A semi-structured interview was chosen as it is an open-ended format, where the interviewee is encouraged to elaborate on their answers in an exploratory way (Dörnyei 2007:136). It was a useful way of getting more information about why they gave the answers that they did on their surveys. The interview questions were based on individual student responses to the two online surveys, if applicable, or questions taken from the online survey if learners had not taken them because they had not been members of the class when the survey was conducted.

I kept a daily journal, where I started by reflecting on the information generated by the online class entry survey. I did the same for the online exit survey. On a daily basis, I described each type of homework task given and the medium it was given in i.e. paper based or online. I recorded my impressions of student reactions to each task, and how I felt learners were reacting to them. In this way I intended to generate data that could be cross-referenced with the online surveys and the interview data.

A qualitative analysis of the interviews, the comments from the online surveys, and the journal was carried out. This involved looking at the interview transcripts, the survey comments and the journal and analysing them for themes. For example, an online student comment about writing online might have corresponded with something they said during the end-of-cycle interview, or an observation that I had made in the journal. I then identified a series of themes that emerged from this analysis. Student attitudes to homework in general, using E Learning and doing traditional types of homework were used as starting points for identifying themes from the data.

With the exception of some of the online survey responses, the scope of the qualitative analysis was restricted to the three learners who had completed all of Cycle 1. A case study approach was adopted for these three learners, because it was the best way of representing the most consistent data from the small number of learners who had completed the cycle.

Findings from Cycle 1

Five of the six learners surveyed stated that they perceived being given homework by the teacher as something of value. In both surveys (Figures 3–4), five out of the six learners disagreed that they should never receive homework, with disagreements becoming slightly stronger in the second survey. Generally, attitudes remained fairly fixed between both surveys. Half the learners were amenable to doing homework with the use of technology,

Figure 3: Cycle 1 entry survey (Total = 6 learners)

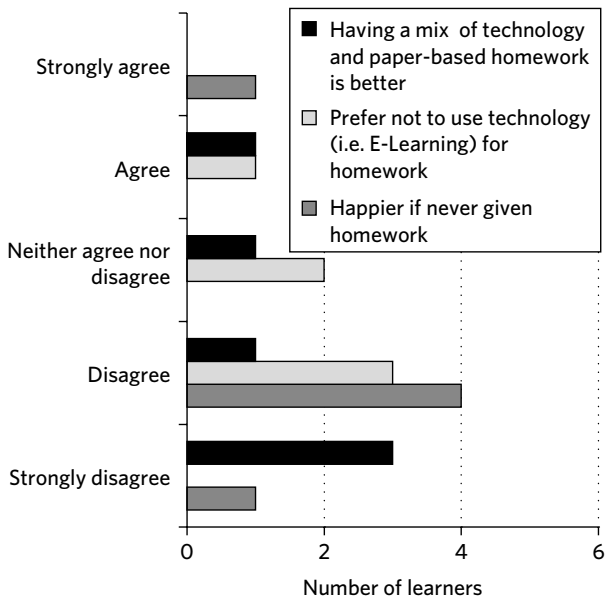


Figure 4: Cycle 1 mid-survey (Total = 6 learners)

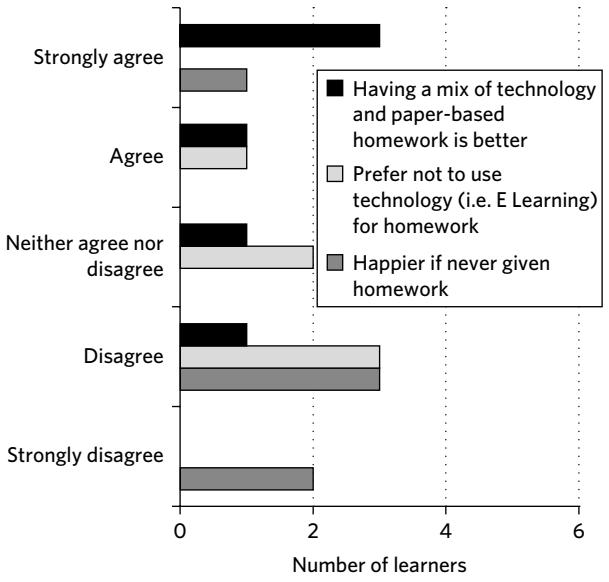


Figure 5: Cycle 1 entry survey (Total = 6 learners)

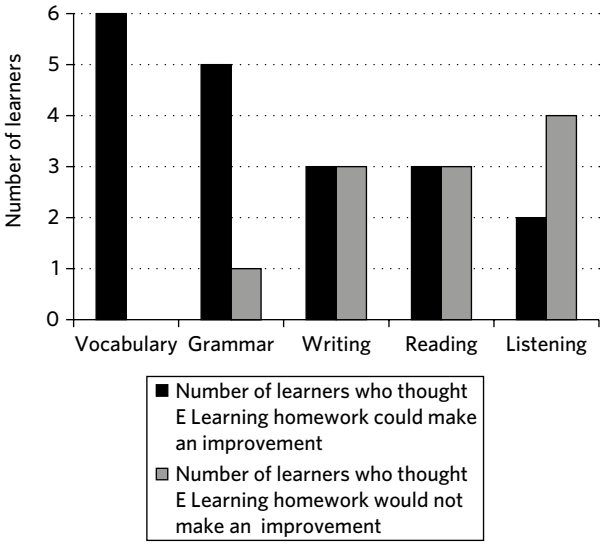
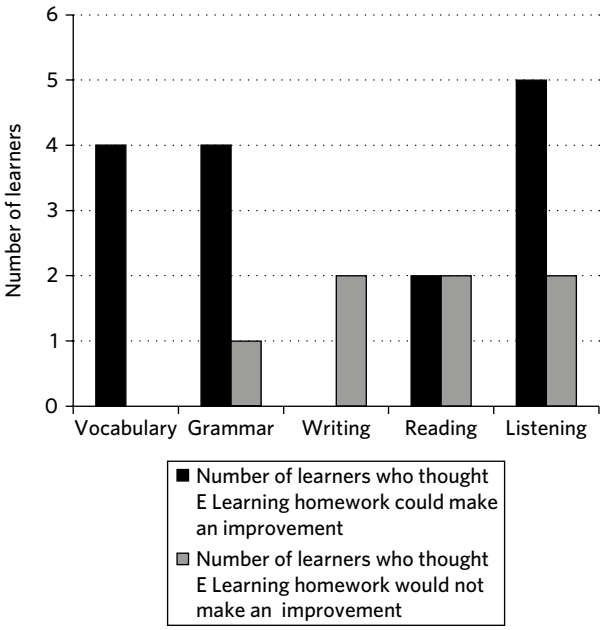


Figure 6: Cycle 1 mid-survey (Total = 6 learners)



while two had no particular preference. One respondent preferred not to use technology (or do any homework). The most notable shift in attitude was towards having a mixture of E Learning and paper-based homework. Between surveys, attitudes became more much favourable with four agreeing that having a mix of paper-based and online homework was better.

Over the first three weeks of this cycle, my learners' perceptions of what could be improved by E Learning saw a more positive shift towards listening (Figures 5–6).

A consistent positive disposition towards using E Learning for grammar study was found, but this attitude was not necessarily reflected in the classroom with other types of homework. During one class I had given my learners a choice between paper-based and online homework:

> For today's homework, for the first time I gave learners the choice between doing paper-based homework and a vocabulary exercise from E Learning . . . two learners had to do the paper-based task, as they are new and don't have their login details yet. However, the majority of the class opted for the paper-based task. (Journal 1, Day 12)

There were clear learner preferences for paper-based homework in Cycle 1, more examples of which will be discussed in the case studies.

There was a small dip in the number of learners who thought reading could be improved by the use of E Learning.

Between the two surveys there was a small drop in the number of students who thought E Learning would improve their vocabulary. The biggest negative change, however, was in writing, where no one saw any benefit to doing online writing. I suspect this is because I only set one online task on Day 18, which was after the mid-point survey. This probably had a negative impact on learner perceptions of the value of using E Learning for writing practice. In Cycle 2, I tried to make sure that we did more online written tasks to get a more balanced view from my learners. In every week of the second cycle I included one main writing homework task, both on E Learning and as a paper-based assignment.

Case study: Leonardo

Leonardo was a 19-year-old Brazilian who had maintained his motivation throughout the cycle, as indicated by the data. His entry and mid-cycle survey results (Appendix 1) indicated a positive attitude towards doing homework, and using technology to study. His preferences for doing grammar tasks came across very clearly on his E Learning study records, which showed that he overwhelmingly used the grammar sections of the website. On Day 4 of the cycle, I had set my class an open-ended task, where learners could choose any section of E Learning. Leonardo's preference for grammar tasks reflected a class-wide perception that studying grammar was more important than studying skills or vocabulary. It does not, however, reflect a preference for doing grammar on E Learning, but for grammar practice *per se*. On Day 6, when class feedback about the homework was done, I noted that:

> Leonardo looked at positive and negative agreement (So do I, Neither do I) . . . Interestingly, despite the latitude given over what learners could choose to study, nearly all chose to study grammar. This reveals more about learner attitudes to what they prioritise as important to study by themselves. (Journal 1, Day 6)

Overall, Leonardo had a preference for studying with paper and pen. He felt that just using E Learning can become repetitive, as the tasks for each section of the website are very similar, as he noted in his interview at the end of the cycle:

> I think you know, because . . . you know for me I am fed up if I do E Learning every time, every day. I need a mix . . . Yeah, every exercise is similar, just change the vocabulary. You need the other types of exercise.

This is obviously a criticism levelled at the school's E Learning platform rather than at online learning in general. 'Other types of exercise' was referring to having more variety of tasks, other than the gap fills and matching exercises that made up most sections of the E Learning website.

It was difficult to assess whether E Learning was having any specific impact on Leonardo's attitudes to doing homework, as he was generally motivated to complete it regardless of the medium. From their study, Wooten and Dillard-Eggers (2013) suggested that learners who have high intrinsic motivation may not need online homework as they can choose any learning method (online or otherwise) that is consistent with their own learning style.

Case study: Sakura

Sakura was a Japanese student in her early 20s who had a strong intrinsic motivation, as measured by her online survey responses. Both online surveys (Appendix 1) remained identical, indicating no discernible change in her attitudes to doing homework online or otherwise. She was a confident user of technology, and valued doing homework and having a mixture of E Learning and paper-based activities.

Sakura's preference for E Learning as a homework tool went against general class preferences for paper-based homework, as noted in my journal entry in the section 'Findings from Cycle 1'.

For today's homework, for the first time I gave learners the choice between doing paper-based homework and a vocabulary exercise from E Learning . . . Two learners had to do the paper-based task, as they are new and don't have their log in details yet. However, the majority of the class opted for the paper-based task. (Journal 1, Day 12)

Sakura liked the grammar explanations in the introductory section of each E Learning study section, as she noted in her interview:

> I like introduction, because they explain a lot of detail . . . So, I like, always I check the introduction. All of sentence.

As with Leonardo, it was difficult to determine whether E Learning had any specific impact on Sakura's attitudes to doing homework, as she already had strong preconceptions about its benefits.

Case study: Yusuf

Yusuf was a learner in his early 20s, and was not interested in studying for a certificate. He was only interested in speaking and listening practice:

> I don't need certificate, and I don't need grammar. After London I have to work with my father and I am working with customer . . . and I have to speak English. I need just speaking and listening.

He did not have negative views about using technology for study, only about doing homework in general, whatever the medium was. His responses on the two online surveys (Appendix 1) remained fairly consistent. However, during his interview in Week 5 of the cycle, he did express a preference for using paper over using E Learning for homework:

> For me, paper is better than E Learning. Because I can see [holds up paper] . . . and paper is better every time.

Yusuf liked the tactile quality that paper had over doing exercises on a screen. He was more amenable to doing

tasks in his free time which would allow him to practise his speaking and listening. In the same interview, he indicated that homework with task authenticity was important for him:

> For example, if you make one group with, you can say that you have to go there . . . And you can give us questions that we have to ask another people.

Reflection on Cycle 1

Overall, there was a class preference towards paper-based homework. My learners also valued grammar tasks more than other types of homework, irrespective of whether it was online or a handout. This suggested that E Learning was not having a significant impact on learner attitudes to doing homework. However, it was not so clear from my research at this point why learners were gravitating towards a paper-based medium. The results of the Cycle 1 research showed a need to have a sharper focus on why my learners were gravitating towards paper-based homework, and away from having a general preference for doing online tasks.

Cycle 2

To further explore learner preferences for paper-based homework, I decided to narrow the research question: To what extent does the homework medium (paper based versus E Learning) impact on learner attitudes to doing homework?

The new research question had a different emphasis from the Cycle 1 question, which was using E Learning as a starting point for investigating impact on learner attitudes to homework. The new question used paper-based and online tasks as a basis for comparing student attitudes. In order to investigate the new research question, I came up with a new research instrument, which was a homework feedback sheet. This was given out with every homework task I set during the second 5-week cycle. On each sheet, the medium in which the assignment was done was recorded (i.e. E Learning or paper based), and students had to indicate a medium preference. They were also invited to comment on what they liked or disliked about each task.

At the end of each week, I recorded the results in a table. For every learner I noted the day the homework was set, a description of the task (i.e. a listening comprehension), the medium (i.e. E Learning or paper), which medium the student preferred for the task, and any comments which they made about their Y/N responses in relation to why they liked or disliked the homework. I also recorded the number of learners out of the class who completed each task. During the cycle I collected 95 homework feedback sheets from my learners, and these provided a good volume of data for analysis.

An average of two E Learning assignments were given per week. In total during the cycle, there were seven paper-based tasks, three occasions where learners had the choice between E Learning or paper-based homework tasks, and nine E Learning tasks. I made sure that two of these were E Learning writing assignments, and I used the E Learning authoring tools to create these tasks for the online platform. This was to address the lack of data about online writing in the first cycle.

In Week 1 and Week 3 my learners completed entry and mid-cycle online surveys. This time the questions were designed to elicit more information about learner medium preferences (Appendix 2). In addition to asking learners about what they thought about having a mix of E Learning and paper-based homework, I asked them how E Learning compared with paper and pen for a number of different task types (e.g. grammar or reading). Learners could choose responses from a Likert scale. Learners were invited to comment on the reasons behind some of these responses in text boxes. Eight learners completed the entry survey and 12 completed the mid-cycle survey. In both cases, the number of students who took the surveys indicated the number of learners who were present in class that day. There was a bigger turnover of learners in this cycle, which meant only four respondents completed both surveys. So this time I decided to use the survey results as case study data only, and not use it to look at general class trends. However, I could not use these same four respondents for the case studies as some were not present at the end of the cycle for being interviewed. This was one of the challenging aspects of doing research with a continuous enrolment class.

For the semi-structured student interviews on the last day of Week 5, I interviewed nine learners. I interviewed the five individually who had completed at least one online survey. My questions were based on their survey responses and comments they had made on their homework feedback sheets. The other four were interviewed in pairs and asked questions from the surveys. All of the interviews conducted were longer than in Cycle 1 (between 5 and 10 minutes) and they were recorded and transcribed.

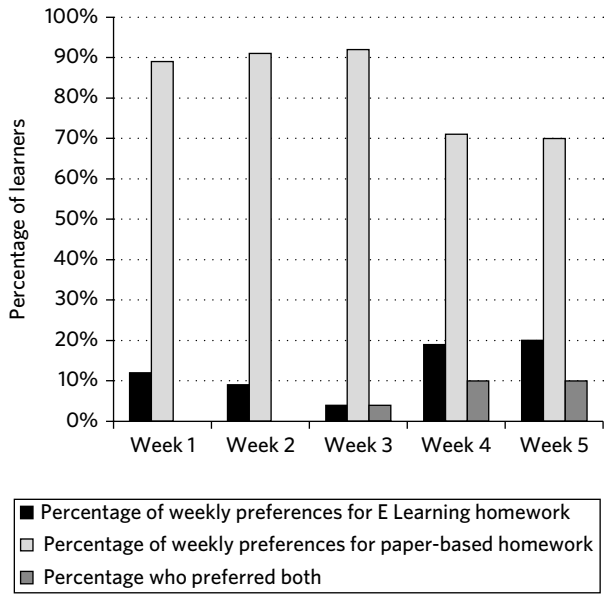
I kept a second journal as a means of using some of the data generated from it for cross referencing with interview and online survey data. It has also been a useful strategy for reflecting on my practice during the period of research.

Data from the homework feedback sheet was analysed quantitatively and presented graphically (Figures 7–10). A qualitative analysis was done on comments from the homework feedback, the learners' interviews and the journal. As in Cycle 1, the sources of data were analysed for themes which could be cross referenced. I chose three learners for case studies. I chose Yusuf again, as he was the only learner who had been in my class throughout the 10 weeks of research. Hana had joined the class just before the start of Cycle 2. I also chose Hiroshi (even though he had started in Week 2 of Cycle 2) as he had provided substantial comments on his homework feedback sheets.

Findings from Cycle 2

Using data from the homework feedback sheets, the total number of weekly homework tasks was counted. The

Figure 7: Weekly preference for E Learning or paper-based homework from feedback sheets



percentage of learners who preferred the task being paper based or both was calculated from each weekly total (i.e. if the homework was an E Learning task, but learners would have preferred it to have been paper based). Figure 7 shows the results.

Overwhelmingly, the evidence shows a strong class bias towards paper-based homework. During Cycle 2, I had noticed that learners were more reluctant to engage with E Learning and had noted this in my journal:

> It really has been a struggle trying to get people to use E Learning, and for various reasons it is being used far less than during the first cycle . . . (Journal 2, Day 17)

Internet connectivity issues could well have influenced class perceptions in a negative way about the feasibility of using the E Learning platform:

> Today I gave everyone the same task – to write a letter of complaint, but I gave them the choice of doing it on the E Learning platform, or doing it on paper. Everyone bar one chose paper! I think this could have been influenced by the class demonstration of how to do the writing task on E Learning. The PC internet connection was very slow, and it took nearly a minute to load up the web page. If this hadn't happened, would more people have chosen the E Learning option? (Journal 2, Day 14)

This may partly be due to the higher turnover of learners during this cycle, as I noted later:

> I am thinking of giving a writing homework from E Learning, but I will also need to think of a paper alternative for those new learners . . . continuous enrolment does make the regular use of the platform a bit unwieldy, as the new learners will be unfamiliar with the system. (Journal 2, Day 22)

In Weeks 4 and 5, there was a slight increase in E Learning preferences, which may be explained by Yusuf's case study.

Figure 8: Weekly percentage of completed homework for E Learning or paper-based mediums from feedback sheets

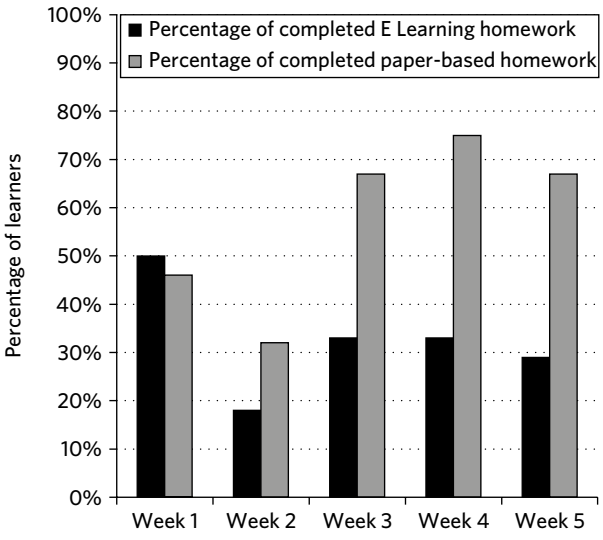
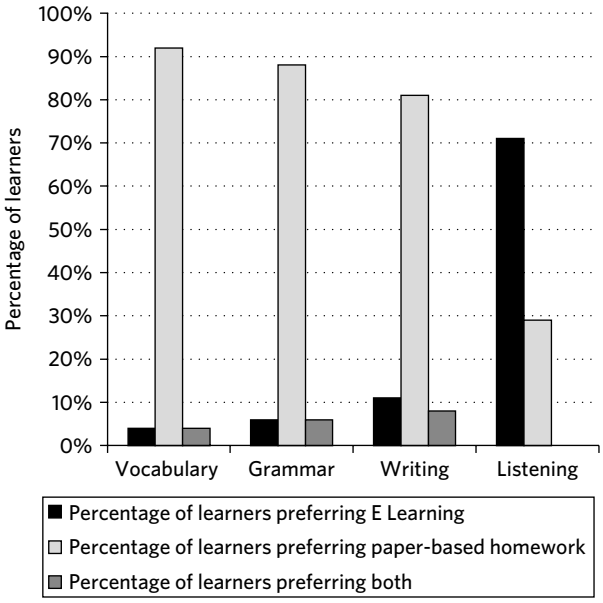


Figure 9: Overall percentage of preferences by task type for E Learning or paper-based homework from feedback sheets



When analysing the percentage of completed weekly E Learning and paper-based assignments (Figure 8), a different pattern emerges. Week 1 shows an almost identical completion rate. The homework set is still successfully completed despite the medium it is done in. After a small dip in Week 2, from Weeks 3 to 5 the percentage of completed E Learning homework remains steady at around 30%, while the percentage of completed paper-based homework is approximately twice the E Learning percentages.

Figure 9 shows the total percentage of learner preferences for E Learning, paper-based homework or both for the indicated tasks for Cycle 2. Again the evidence is very strongly weighted against E Learning for all tasks with the exception of listening, and a small increase for writing.

Comments from the entry survey also corroborate these preference patterns:

> Some homework I need to paper but listening homework must use computer.

> For writing I prefer use pen and paper, but for listening online will be much better.

In the mid-cycle surveys, some comments focused more on technical issues, reflecting some of my concerns about internet connection speeds above:

> I can't use a computer in the cafeteria, either because of it was bad connection.

> Because sometimes connection Wi-Fi doesn't work.

In addition to strong overall indicated preferences for paper-based homework in Figure 9, there seemed to be the feeling from some learners that paper-based homework was a better medium for doing homework in, as can be seen from Hana's case study below.

Case study: Hana

Hana was a Korean student in her mid-20s, who joined the class before the end of the first cycle. In her entry and mid-Cycle 2 surveys (Appendix 2), she indicated a preference for listening to be done on E Learning as she could repeat the audio as often as she liked:

> Because, just one time listening I don't understand is faster listening, and I want more, more, more! Again, again listening.

On her homework feedback sheet, the only time she expressed a preference for E Learning was for listening. She was overwhelmingly in favour of paper-based homework. She said that using paper helped her to concentrate when she was studying, and that using a computer screen could be distracting.

Hana: Yes, I look just the screen, I don't concentrate . . . Not focus.
Teacher: You can't focus, yeah.
Hana: My mind is literally is I can search another thing and I can't . . .
Teacher: You can focus better with paper.
Hana: Yes.

Like Yusuf from Cycle 1, she prefers the more tactile nature of paper, to the use of a computer screen. As a physical record of what she studied, it helps her to remember what she has learned.

Hana: I think touching the paper is more remember, more easily remember.
Teacher: It helps you to remember more easily, yeah?
Hana: Yes.

She found the design of E Learning writing tasks to be off-putting. On E Learning, the recipient of the graded writing gets both the original text box that contains the student's work and a second text box with the corrected teacher's version. Hana found the spatial separation of original and corrected work difficult to process.

Hana: Very difficult is compare . . .
Teacher: Because it's not in the same place?
Hana: Yeah . . . I just see the is whole thing, but E Learning writing practice is my wrong sentence here [one text box] is your correct sentence here [another text box] . . .

With the exception of listening tasks, Hana preferred using paper with nearly all types of tasks, both as a physical record of her work and as a means of helping her study more effectively. It is worth noting that her negative response to doing writing tasks on E Learning could well have been due to the limitations of the E Learning platform as opposed to being a negative response to online study generally.

Case study: Hiroshi

Hiroshi had a more open-minded approach to the medium in which he did his homework. On his homework feedback sheets, his preferences were weighted equally – four for E Learning, four for paper and four for both. In his mid-cycle survey comments (Appendix 2), he did not state a particular preference for either paper-based or E Learning homework.

> I have no preference between paper and pen and E Learning: it depend on the type of study.

In his interview, he emphasised the importance of the nature of the homework task over the medium. With grammar practice, having a substantial number of exercises which were repeating the target language was more beneficial for him, as he associated it with forming linguistic habits:

Hiroshi: I think that practice for grammar skills is not important on E Learning or with pens . . . And I think is more important thing is . . . the amount of questions, so how many numbers or –
Teacher: Oh, the number of exercises?
Hiroshi: Yeah . . . about grammar.
Teacher: About grammar?
Hiroshi: Yeah, yeah. Grammar training is maybe on, hmm, how many I tried. Or, yes –
Teacher: So if you have lots of exercises, you can just repeat, repeat, repeat?
Hiroshi: Yeah.
Teacher: It gives you muscle memory.
Hiroshi: Yeah, that's right. Yeah, it's better, so it's not important the way, E Learning or paper.

Hiroshi's interview comments about study and repetition above were also consistent with his homework feedback comment about practising gerunds and infinitives, although he tentatively opted for E Learning as a better medium for the study of verb patterns:

> In this type of training, it's better to deal with a large number, I think. So doing this on E Learning is better . . .

Using a paper medium was his preferred medium for grammar practice, but Hiroshi also concedes that instant feedback when doing an online grammar task could also be helpful. In one of his homework feedback comments about a grammar gap fill, he lists both the positive and negative aspects of doing this on paper:

> (+) Writing with paper and pen is good training to understand grammar.
>
> (−) If we had done this on E Learning, we could have known the answer immediately.

Like most of the class, Hiroshi preferred a paper-based medium for writing practice, as he found it easier to notice his mistakes when writing with a pen as opposed to typing out an exercise:

> This is a practical training for when we'll have to explain about something. I think it's easier to find miss spellings [sic] with paper and pen. Actually, I noticed many miss spellings [sic] myself. (Day 18, homework feedback sheet)

During the interview, he reinforced this view:

> Yeah . . . I think that . . . writing on paper with pen is more easy to find my mistakes . . . I write wrong spelling . . . Maybe on the internet I didn't notice my misspellings. But when I write on the paper, I think I can recognise my misspellings.

However, Hiroshi also acknowledged both mediums (online and paper based) are good for writing practice. He found a CV writing exercise much easier when he switched from paper to E Learning. On his homework feedback sheet, he wrote:

> I tried to make a CV with paper and pen, but it was difficult a little bit. So I changed my plan and did homework on E Learning. It was also difficult but a little bit easier because it was a form of 'fill in the gap'. I think both of them are good for training. (Day 23, homework feedback sheet)

In this case I think this is more a comment about the nature of the task, rather than the medium it was done in. The paper-based exercise was more unguided, and the E Learning task was much more guided, as it was a gap-fill activity.

Case study: Yusuf (Cycle 2)

In Cycle 1, Yusuf had shown a lot of resistance towards doing most kinds of homework, whatever the medium. However, between his Cycle 2 entry and mid-cycle surveys (Appendix 2), his views on having a mix of paper-based and online assignments changed greatly. Although attitudes to doing homework had remained somewhat the same, his attitude to using E Learning for writing practice was the most marked change. His homework feedback comment about writing a guide to his city on E Learning (Day 18) was his first positive one about the medium.

> It was easy I can check my writing part in english.

I investigated this more during his second interview:

> Yeah, city guide and you can describe your country and E Learning you don't need spend your effort for writing and completely easier than writing. It was good.

Soon after he had completed this task, I noticed that Yusuf had again opted for an E Learning assignment. I had given the class a choice between doing the homework on E Learning or on paper.

> Interestingly enough, Yusuf chose the E Learning task! . . . He seems to have changed his mind about using the E Learning platform for writing, at least. Most others opted for finishing their written CV [on paper]. (Journal 2, Day 23)

During his interview, I asked him why he had chosen E Learning.

> Because I was feeling like I don't want to write, I don't want to spend my effort. E Learning is easier than writing. Because you can [type] something on the computer . . . and paper and pen, I think you have to spend your time more than E Learning.

Yusuf was reluctant to spend too much study time on homework, but found the online medium was more suitable for his learning style than doing writing practice on paper.

Discussion

The first cycle of this project focused on the impact of E Learning on learner attitudes to homework. The general trend seems to be that there were no clear preferences for E Learning or other online homework platforms. Listening via E Learning was seen as beneficial, with most learners agreeing that a mixture of online and paper-based homework is better. However, by the end of the cycle, there had been a more marked preference for paper-based homework, with E Learning having little impact on attitudes.

The second cycle saw a much stronger bias towards paper-based homework, with the exception of listening practice. Writing task design may have been one of the contributory factors. As discussed, there were also internet connectivity problems which may have led to some perceptions of E Learning being impractical to use. There were also E Learning training issues with having a continuous enrolment class. Unlike the study by Doorn et al (2010), my research did not find a stronger preference for online homework.

The case studies impressed upon me the fact that the homework, irrespective of the medium, needed to suit individual learner needs. With some of the learners (Leonardo and Hiroshi) we saw that the focus of the task (i.e. grammar) was more important than the medium in which it was completed. This is corroborated by homework research conducted outside the EFL sector, which suggests that individualised tasks (not the medium in which they were completed) can be more meaningful, engaging and motivating for learners (Darling-Hammond and Ifill-Lynch

2006, National Education Association no date, Sharp, Keys and Benefield 2001).

Having a variety of homework tasks may have a more significant impact on learner preferences than the medium in which the tasks are done. Sharp et al (2001) point out that their research into learner preferences showed setting routine or repetitive assignments does not contribute to learning. They also acknowledge that research into student preferences is limited and that there is an urgent need for further research into the impact of homework on learning attitudes and the application of new technology to homework.

Willingness to complete homework tasks seemed to be linked to motivation, in addition to being connected to individual learner needs. Leonardo and Sakura had a positive disposition towards using E Learning for studying[1], but would have been generally happy using any medium because they were both highly motivated, as Wooten and Dillard-Eggers (2013) had said about motivated learners in their study. Hiroshi would judge each homework task on its own merits, irrespective of the medium. Hana preferred paper-based homework as it suited her learning style more. She was put off by the visual design of the E Learning website, which implies that she had problems responding to the platform, rather than having a negative attitude to studying online generally. Yusuf had been fairly resistant to doing homework, but discovered later on that he preferred using E Learning for writing practice, because it was better for his studying style. Both writing tasks that he completed had a high relevancy to his own life. These tasks could have reasonable claims to being functionally authentic as argued by Buendgens-Kosten (2013), as well as having task authenticity as argued in the literature above.

The overall trend seemed to indicate that the medium was less important to learners than other factors that influenced how they felt about the particular task that they were doing.

Limitations

One of the biggest challenges was doing a longitudinal study with a continuous enrolment class. Doorn et al (2010) and Wooten and Dillard-Eggers (2013) studied closed classes in a university setting. As such they could study classes with the same learners over a period of many weeks. Their learners all had common goals (i.e. to pass a course), whereas in my General English class my learners all had different goals, and the composition of the class was changing every week.

My research had a very local context, covering a small number of learners and one teacher's impressions. As such it is worth noting that my learners may have responded differently to another online platform, where the school had not invited all of my learners to use a common online platform that had been developed and purchased by the institution I work for.

As a result of carrying out this research, it has become more apparent to me that I cannot make assumptions that learners will automatically find E Learning more appealing than paper-based tasks. By extension, this also has implications for educational organisations, which should avoid the danger of assuming that anything online is better or more effective than using traditional media, which is clearly not the case from my experience.

Future research

My research has made me realise how doing an analysis of my learners' needs should be an integral part of what homework tasks I set them to do, whatever the medium of the homework is. This is much more challenging to do in a General English continuous enrolment class, where individual learning goals are diverse and class dynamics are constantly changing.

The regular needs analysis that I do with my learners needs to ask about the types of homework tasks they would find beneficial and what they would like to learn outside the classroom as well as in lessons. It implies setting different tasks for different learners (whether online or paper based), which presents another set of practical challenges for full-time teachers like myself. If there was a Cycle 3, I would widen the scope of enquiry. Would learners from other classes have the same overwhelming preferences for paper-based homework? Were these findings unique to my own class and teaching context or could there be something more generally applicable about them? Although what I have learned is connected to my own teaching practice, there are some implications for the school where I work. As part of the school orientation for new learners, some kind of practical E Learning induction would lessen the need for training to be done in classes, which might make learners more amenable to using the platform. On reflection, perhaps making a distinction between paper-based and online ways of doing homework was not the best thing for my learners. What is relevant for my learners exceeds considerations of what medium the homework is completed in.

Reflections

Through the process of doing action research, I was able to allay my initial anxieties about getting enough consistent data in a class of constant changes. Choosing a case study format for presenting most of my findings was the most practical way of addressing this concern. Keeping an open mind, and letting the process of the research inform the direction it takes for the next cycle, is one of the key things I learned from doing action research.

[1] With the exception of Leonardo's negative feedback about the repetitive nature of some tasks (Appendix 1).

I learned from Cycle 1 and then improved upon my research design in Cycle 2. Some of the questions I used in the entry survey (Cycle 1) were different for the mid-cycle survey ones, and I quickly learned this would not help me observe potential changes in attitudes over time. I was able to improve upon this approach for Cycle 2. The study needed a narrower focus to examine learner preferences, and after revising my research question, I quickly discovered that I would need to generate more quantitative data in Cycle 2 about the homework tasks that were completed.

Being involved in this project challenged my assumptions about my learners and how they feel about doing online homework. It was a process of adopting a critical or questioning stance towards my own teaching practice (Burns 2010). Doing action research has helped me to develop as a researcher and as a practitioner. As such, the combination of two processes I had previously thought of as being independent from each other (doing research and developing as a teacher), proved to be an effective and engaging way of learning more about my own teaching. It was also a rewarding way of going into more depth with learner needs than I would have otherwise done. In the future, I would be very interested in applying action research techniques to other aspects of teaching, or even collaborating with colleagues to explore methods and approaches as a community of practice.

References

Buendgens-Kosten, J (2013) Authenticity in CALL: Three domains of 'realness', *RECALL* 25 (2), 272–285.

Burns, A (2010) *Doing Action Research in English Language Teaching: A Guide for Practitioners*, New York: Routledge.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.

Darling-Hammond, L and Ifill-Lynch, O (2006) If they'd only do their work! *Educational Leadership* 63 (5), 8–13.

Doorn, D, Janssen, S and O'Brien, M (2010) Student attitudes and approaches to online homework, *International Journal for the Scholarship of Teaching and Learning* 4 (1), 1–19.

Dörnyei, Z (2007) *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*, Oxford: Oxford University Press.

Hong, E, Wan, M and Peng, Y (2011) Discrepancies between learners' and teachers' perceptions of homework, *Journal of Advanced Academics* 22 (2), 280–308.

National Education Association (no date) *Research Spotlight on Homework: NEA Reviews of the Research on Best Practices in Education*, available online: www.nea.org/tools/16938.htm

North, S and Pillay, H (2002) Homework: Re-examining the routine, *ELT Journal* 56 (2), 137–145.

Painter, L (2003) *Homework*, Oxford: Oxford University Press.

Sharp, C, Keys, W and Benefield, P (2001) *Homework: A Review of Recent Research*, available online: www.nfer.ac.uk/publications/HWK01/HWK01.pdf

Wooten, T and Dillard-Eggers, J (2013) An investigation of online homework: Required or not required? *Contemporary Issues In Education Research* 6 (2), 189–198.

Appendix 1: Cycle 1 sample survey responses

	Leonardo		Sakura		Yusuf	
	Entry survey	Mid-cycle survey	Entry survey	Mid-cycle survey	Entry survey	Mid-cycle survey
I think that using technology (computers, smart phones, tablets etc.) helps me to learn English.	Agree	N/A	Agree strongly	N/A	Agree	N/A
I am a confident user of technology (computers, smart phones, tablets etc.).	Agree	N/A	Agree	N/A	Agree	N/A
I enjoy using technology (computers, smart phones, tablets etc.) to help me with my studies.	Agree strongly	N/A	Agree strongly	N/A	Agree strongly	N/A
How much of this time using technology (computers, smart phones, tablets etc.) do you spend on studying English?	2–3 hours	1–2 hours	30 mins to 1 hour	30 mins to 1 hour	2–3 hours	Less than 30 mins
I prefer to do homework where I do not need to use any technology (computers, smart phones, tablets etc.).	Disagree	Neither agree or disagree	Neither agree or disagree	Neither agree or disagree	Agree	Disagree strongly
I think it is better to have a mix of technology (computers, smart phones, tablets etc.) and traditional ways of doing homework (e.g. paper and pen).	Agree	Neither agree or disagree	Agree strongly	Agree strongly	Disagree	Disagree strongly
Both studying with the teacher in class and doing homework in my own time is important.	Agree strongly	Agree strongly	Agree strongly	Agree strongly	Disagree	Disagree
How often do you think your teacher should give you homework?	Every day	Every day	Every day	Every day	Never	Once a week
I would be happy if my teacher never gave me homework.	Disagree strongly	Disagree strongly	Disagree	Disagree strongly	Agree strongly	Agree strongly
Which of the following do you think you can improve by using E Learning to do your homework? (You can choose more than one.)	Writing, vocabulary grammar	Listening, vocabulary, grammar	Grammar, vocabulary, listening, reading	Grammar, vocabulary, listening, reading	Vocabulary, Listening, Speaking	Listening, Speaking
What would you enjoy doing for homework? For example, watching the news or writing a story.	Complete the gaps and writing.	Listen the music, complete the gaps.	Write an essay of newspaper.	Watching the news or sitcom. Writing essay about newspaper.	We have to talk about something	I don't want homework we have to speak with peoples its better than homework.

Appendix 2: Cycle 2 sample survey responses

	Hana		Hiroshi		Yusuf	
	Entry survey	Mid-cycle survey	Entry survey	Mid-cycle survey	Entry survey	Mid-cycle survey
How often do you use E Learning?	1 to 2 times a week	1 to 2 times a week	N/A	1 to 2 times a week	Never	1 to 2 times a week
I like using technology (computers, smart phones, tablets etc.) more than using paper to do homework.	Neither agree nor disagree	Neither agree nor disagree	N/A	Neither agree nor disagree	Strongly disagree	Neither agree nor disagree
I prefer to do homework where I do not need to use any technology (computers, smart phones, tablets etc.).	Neither agree nor disagree	Neither agree nor disagree	N/A	Agree	Disagree	Neither agree nor disagree
I think it is better to have a mix of studying online (e.g. Using E Learning) and have traditional ways of doing homework (e.g. typing or writing).	Strongly agree: some homework i need to paper but listening homework must use computer.	Strongly agree: Sometimes paper homework better than online homework.	N/A	Strongly agree: It depend on the type of study.	Disagree	Agree: its good way for study.

	Hana		Hiroshi		Yusuf	
	Entry survey	**Mid-cycle survey**	**Entry survey**	**Mid-cycle survey**	**Entry survey**	**Mid-cycle survey**
How does using E Learning to do homework compare with using paper and pen to do homework?	Using pen and paper is better: i think the best way is mix!!!! but i prefer paper. but e-learning is no problem.	Using paper and pen is better: I prefer paper homework, but sometimes listing and grammar homework better than paper.		I have no preference between paper and pen and E Learning: it depend on the type of study.	Using paper & pen is much better: i don't like using computer and something.	Using paper and pen is better: paper and pen more useful.
Which of the following is better as online homework (E Learning) and which is better as traditional homework (typing or writing)?	E Learning: Listening, Pronunciation, Speaking, Spelling, Vocabulary. Traditional: Reading, Writing.	E Learning: Listening, Pronunciation, Speaking. Traditional: Reading, Spelling, Vocabulary, Writing.		E Learning: Listening, Pronunciation, Speaking. Traditional: Spelling, Vocabulary, Writing. No difference: Reading.	It makes no difference: Listening, Speaking, Vocabulary.	E Learning: Listening, Reading, Speaking, Vocabulary.
I like using the internet (e.g. E Learning) for extra self-study and not just for doing homework my teacher gives me.	E Learning: Listening, Pronunciation, Speaking, Spelling Vocabulary.	Agree: I think it very useful for my english study.		Agree: I can choose contents which I want to try at the moment.	Disagree: if you ask me when you are talking with people its much better homework or e-learning because i just want to speaking and listening that why i want chose that.	Neither agree nor disagree: I have no idea.

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Studies in Language Testing

An indispensable resource for anyone interested in new developments and research in language testing

To find out more about our full list of publications:

www.cambridge.org/elt/silt

www.cambridgeenglish.org/silt

To subscribe to *Research Notes* and download previous issues, please visit:
www.cambridgeenglish.org/research-notes

Contents

For further information visit the website:
www.cambridgeenglish.org

Cambridge English
Language Assessment
1 Hills Road
Cambridge
CB1 2EU
United Kingdom

www.cambridgeenglish.org/helpdesk

A DIVISION OF CAMBRIDGE ASSESSMENT

ALTE

THE QUEEN'S AWARDS FOR ENTERPRISE: 2015

0448235703

All details are correct at the time of going to print in October 2016