CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

100 CAMBRIDGE ENGLISH
CENTENARY 1913–2013

Research Notes

Issue 54

November 2013

ISSN 1756-509X

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Research Notes

Issue 54 / November 2013
A quarterly publication reporting on learning, teaching and assessment

Senior Editor and Editor
Dr Hanan Khalifa, *Head of Research and International Development*, Cambridge English Language Assessment
Coreen Docherty, *Senior Research and Validation Manager*, Cambridge English Language Assessment

Editorial Board
Dr Fiona Barker, *Senior Research and Validation Manager*, Cambridge English Language Assessment
Clare Harrison, *Assessment Group Manager*, Cambridge English Language Assessment
Dr Nick Saville, *Director*, Cambridge English Language Assessment
Lynne Stevenson, *Assessment Group Manager*, Cambridge English Language Assessment

Production Team
Rachel Rudge, *Production Controller*, Cambridge English Language Assessment
John Savage, *Publications Assistant*, Cambridge English Language Assessment

Printed in the United Kingdom by Canon Business Services

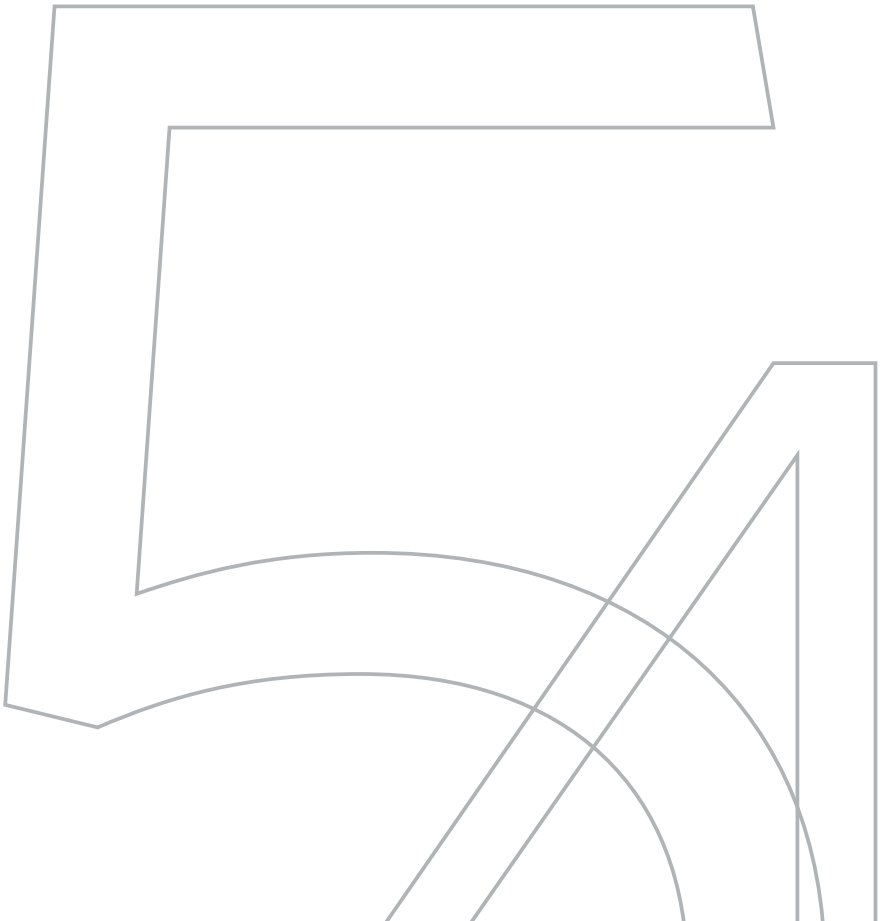
Research Notes

Issue 54

November 2013

Contents

Editorial	**2**
Delta trainers' cognitions and pedagogical practices Simon Borg	**3**
Washback effect of Cambridge English: Key for Schools in São Paulo, Brazil: A case study Matilde Virgínia Ricardi Scaramucci and Eliana Kobayashi	**13**
Young learner perceptions of test taking in Japan Jennifer M Toews-Shimizu	**21**
Stakeholders' perceptions of the Occupational English Test (OET): An exploratory study Ivana Vidaković and Hanan Khalifa	**29**
Book announcement: Validity in Educational and Psychological Assessment Paul Newton and Stuart Shaw	**33**
Caroline Clapham IELTS Masters Award 2013	**36**



Editorial

Welcome to issue 54 of *Research Notes*, our quarterly publication reporting on research matters relating to learning, teaching and assessment within Cambridge English Language Assessment. This issue primarily focuses on the research outcomes from the third round of the Cambridge English Funded Research Programme.

The Cambridge English Funded Research Programme, was launched in 2009 to make funding available for small, medium and large-scale research projects that focus on our exams and issues related to test validation, contexts of test use, test impact or learning-oriented assessment. The Programme represents the commitment of Cambridge English Language Assessment not only to ensure our tests are fair, accurate and valid but also to support educational professionals who are at different stages of their research career. The researchers showcased in this issue have all taken a case study approach to investigate the complexities of learning, teaching and assessment.

The first article, by Simon Borg, focuses on the beliefs that underpin Delta (Diploma in Teaching English to Speakers of Other Languages) trainers' pedagogical practices. Borg draws on video-recorded input sessions (i.e. seminars, lectures and workshops led by Delta trainers) and follow-up interviews to investigate trainer principles and practices. He found that there was a link between trainers' practices and their core beliefs even though there was some variation in terms of their ability to articulate these beliefs. Borg's findings are an important contribution to the under-researched field of language teacher education.

The next article by Matilde Virgínia Ricardi Scaramucci and Eliana Kobayashi focuses on the washback effect of *Cambridge English: Key for Schools* in Brazil. This case study uses interviews, classroom observations and a review of school documents to come to an understanding of the effects of using this exam on teaching practices. Scaramucci and Kobayashi found that the *Cambridge English: Key for Schools* exam is in line with and supports the school curriculum which emphasises communicative language instruction. As a result, the constructs and content of the exam do not seem to be overtly influencing the choice of textbook or teaching practices but rather the exam seems to complement both. They did find, however, that the multiple-choice task type found on the exam did influence the design of their achievement tests as this particular task type is one that was not typically used before Cambridge English exams were introduced into the curriculum. This case study demonstrates how schools can integrate Cambridge English exams into a curriculum in a positive manner which supports language learning.

The last article from the Cambridge English Funded Research Programme is by Jennifer Toews-Shimizu, who reports on Japanese young learners' perceptions of the *Cambridge English: Starters* exam. The aim of her study was to investigate whether test design could have a positive impact on learner attitude and confidence by analysing written reflections produced by students pre- and post-exam. Her findings suggest that students who are highly engaged and performing well in English class tend to be more confident and less anxious about the test than students who are less engaged and not performing as well in class. This study not only provides an insight into Japanese young learner attitudes to tests but also highlights an important issue about the use of tests; in particular, using tests for purposes for which they weren't necessarily originally intended. In order to ensure her students took the testing event seriously, Toews-Shimizu graded her students' performance on the test and incorporated these scores into their class grades, which ultimately changed the intent and stakes of the exam. The *Cambridge English: Young Learners* suite of exams were designed to offer learners a positive first experience of international tests. There is no pass or fail distinction as achievement is rewarded with 'shields'. One could argue that assigning scores that affect overall grades could undermine the intended purpose of the exam, resulting in negative impact by increasing test taker anxiety. Although her findings suggest that this did not happen, repeated or prolonged misuse of a test may have just such an effect, raising questions about the extent to which test developers are responsible for or can control how their tests are used. Inevitably, they can't, which is why it is important for them to provide stakeholders with comprehensive and transparent information about their tests and their intended purposes in order to encourage users to reflect on whether a particular test is appropriate for a particular context.

The next two articles explore two important issues: the language skills of healthcare workers and the meaning of validity. The Occupational English Test (OET) was designed for professionals in 12 health-related fields who need to demonstrate they have the language skills necessary to work in an English-speaking environment. Vidaković and Khalifa investigated the impact of this test by surveying former test takers, teachers who prepare students for this exam, employers and regulatory bodies about the appropriateness of the test. The authors found that stakeholders have positive attitudes towards the test because the content and tasks are authentic. In addition, candidates report that studying for the test increased their confidence in doing their job.

The last article in this issue describes the content of a forthcoming book titled *Validity in Educational and Psychological Assessment* by Paul Newton and Stuart Shaw. The book, which will be published by SAGE in early 2014, explores what it means to make a validity claim and how the concept of validity has changed over time. In the book, the authors will provide a framework for evaluating tests which expands on Messick's 'progressive matrix'. Finally, we are pleased to announce the 2013 Caroline Clapham IELTS Masters Award.

Hanan Khalifa
Senior Editor

Coreen Docherty
Editor

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Delta trainers' cognitions and pedagogical practices

SIMON BORG SCHOOL OF EDUCATION, UNIVERSITY OF LEEDS, UK

Introduction

It is widely accepted (e.g. Borg 2006, Johnson and Golombek (Eds) 2011) that language teacher development cannot be effectively promoted without an understanding of the cognitions – i.e. beliefs and knowledge – that underpin what teachers do. Yet while the same principle applies to teacher educators, empirical interest in language teacher educators' work – in what they do and why – has been, as noted by Borg (2011), very limited. In contrast, the broader educational literature has increasingly recognised the study of teacher educators as an important focus for research (see, for example, Bates, Swennen and Jones (Eds) (2011) and special themed issues of *Teaching and Teacher Education* (21 (2), 2005) and the *Journal of Education for Teaching* (37 (3), 2011)). Koster, Brekelmans, Korthagen and Wubbels also argue that 'increasing the knowledge of tasks and competencies of teacher education contributes towards strengthening the position of the profession and the possibilities for professional development' (2005:158). The lack of such knowledge in the field of language teacher education seriously constrains our ability to make informed proposals about how the development of language teacher educators can be effectively supported.

In response to this gap, I examined the pedagogical practices and cognitions of seven language teacher trainers, and this report presents an overview of some key findings from this work and illustrates them with case studies of two of the participating trainers.

Context

Participants in this study were all trainers on Cambridge English Language Assessment's Diploma in Teaching English to Speakers of Other Languages (Delta) teaching award. This is an internationally recognised advanced course for practising teachers and consists of three modules: 1) Understanding Language, Methodology and Resources for Teaching, 2) Developing Professional Practice and 3) Extending Practice and ELT Specialism. Different models of course delivery exist (e.g. full time, part time and blended learning). Modules can be studied and certified individually and on completion of all three the Delta is awarded. Preparation courses are offered for each module concurrently; however, attendance on a course is only necessary for Module Two. Full details of the objectives of each module and of the Delta syllabus and assessments are available online: www.cambridgeenglish.org/exams-and-qualifications/delta/

Cambridge English provides the *Training and Induction Handbook* (Cambridge ESOL 2009) which outlines the requirements Delta trainers must meet. According to this document, all prospective Delta trainers should have substantial English Language Teaching (ELT) experience, hold a Delta or equivalent, and have demonstrated a commitment to their own professional development. Delta trainers-in-training will go through one or more induction, training and development programme depending on the degree and nature of their existing experience as a teacher educator. On conclusion of this process they must submit a portfolio which will include a range of evidence, again dependent on the kind of preparatory programmes they have been through. For example, prospective trainers will include plans and materials for any input sessions they have given and feedback from other tutors who observed these sessions. The Handbook specifies a list of competencies which potential trainers need to meet. One of these is that they must be able to 'design and deliver input sessions at an appropriate level and in relation to the Delta syllabus areas' (Cambridge ESOL 2009:10). Input sessions are particularly relevant to Delta Modules One and Two, as Module Three is largely an independent research project. Input sessions include seminars, lectures, and workshops that Delta trainers lead. They do not include supervised teaching or post-observation feedback sessions.

Research questions

This study addressed the following questions in relation to the work of Delta trainers:

a) What pedagogical practices characterise the way trainers deliver input sessions?

b) What principles underpin these practices?

c) What influences have shaped the development of a) and b) above?

Methodology

Participants

Cambridge English assisted with recruitment by inviting Delta trainers to participate in this study via their network. Of the 39 trainers who initially expressed an interest, the final sample was reduced to seven after further discussions between the researcher and participants about the scope of the study. Table 1 summarises the trainers' profiles.

Data collection and analysis

Table 2 summarises the chronology of data collection with each of the seven trainers. All participants were first sent a background questionnaire (BQ). This then formed the basis of the first interview with each trainer, which (as with all except two interviews in this study[1]) took place by telephone and was, with permission, audio-recorded. The purpose of

[1] Two interviews with one of the trainers in Turkey were conducted face to face as I happened to visit their institution.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Table 1: Participating trainers' profiles

Trainer	Region	ELT (years)	Delta trainer (years)	Qualifications
TR1	Australasia	20+	16–20	Certificate, Diploma, MA
TR2	Turkey	20+	11–15	Certificate, Diploma, BA, MA
TR3	UK	20+	1–5	Certificate, Diploma, MA
TR4	Middle East	20+	1–5	Certificate, Diploma, MA
TR5	Turkey	20+	11–15	BA, MA, PhD
TR6	UK	20+	20+	Certificate, Diploma, BA
TR7	Middle East	20+	11–15	Certificate, Diploma, MA, PhD

Table 2: Data collection

Trainer	BQ	VS1	S2	S3	I1	I2	I3
TR1	8/2	✓	✓	✓	28/2	25/4	2/5
TR2	23/2	✓	✓	✓	2/3	26/4	28/5
TR3	14/2	✓	✓	✓	24/2	2/5	16/5
TR4	12/2	✓	✓	✓	2/3	12/5	16/6
TR5	29/2	✓	✓	✓	23/3	19/6	20/6
TR6	23/2	✓	✓	✓	13/4	15/5	15/6
TR7	1/4	✓	✓	✓	22/4	27/6	22/10

[Key: BQ = background questionnaire; VS1 = video-recorded input session; S2/S3 = materials for input sessions 2 and 3; I1/I2/I3 = the three interviews]

this interview was to discuss trainers' entry into the field of English as a Foreign Language (EFL) generally and teacher training specifically, and to trace key influences on their development as trainers over the years. Following the first interview, trainers were asked to submit a video of one of their input sessions plus a 'lesson plan' and supporting materials for at least three sessions, including the one that was video-recorded. My analysis of these materials (i.e. identifying the kinds of training activities and processes used) provided the basis for the second and third interviews with the trainers. The purpose of these interviews was to explore how trainers deliver input sessions and why. All interviews lasted approximately 60 minutes and were fully transcribed.

In line with the qualitative principles guiding this study, data analysis took place concurrently with data collection. For example, the BQ was analysed to provide the basis for the first interview, which was analysed in preparation for the second, as were the input session materials provided; and the second interview was analysed in preparation for the third.

Findings

Given the volume of data generated by the study and limited space available here I will present a detailed analysis of two of the cases followed by an overall discussion of the key findings to emerge from a consideration of all seven cases.

Case study 1: Victoria (TR2)

Victoria's parents were schoolteachers in the UK and she noted early in our first interview that *I'd always said I wouldn't be a teacher* (TR2:I1). She had in fact been in EFL for over 20 years at the time of this study, spending her entire career in Turkey. She had finished her Certificate in Teaching English to Speakers of Other Languages (CELTA) in 1990 (in the UK) and Delta in 1993 (in Turkey). She was made Head Teacher in her very first job after completing the CELTA and therefore had responsibility for supporting staff and running training workshops from the outset of her career in EFL. She had been a Delta trainer for 12 years.

Features of input sessions

The three input sessions Victoria submitted for analysis were: The articles (Module One, 60 mins), Lexis (Module Two, 180 mins), and Second language acquisition (SLA) (Module One, 60 mins). Each of these sessions was taught with a group of around 25 Delta trainees and an analysis of them highlighted the following features of Victoria's practices during input sessions.

1. Trainee-centred tasks (e.g. listing, sorting, gap-filling, matching) that require trainees to work in groups, interact and discuss (e.g. the session on articles, where groups analyse sentences, work out rules then crossover and discuss).
2. Loop input, where this can be achieved (e.g. the session on lexis – teaching about teaching vocabulary by teaching vocabulary).
3. Handouts/worksheets, typically containing interactive tasks or 'answer keys' (see below). She explained that *I like materials-heavy and centre-of-focus-teacher-light teaching* (TR2:I1) and elaborated on the role of worksheets as follows:

   The worksheets are a way of feeding things in, but getting them to work through, so they have to, either one of them knows and the other one doesn't, and they sometimes have to make gallery readings, or information, or jigsaw or sometimes it's self-discovery. But in a way, that's a way of getting stuff in that I think that some of them don't know that I think they need to know. Now all of it they're going to have to go on and read more about it, but at least it's a taster for them. So that's that bit of it delivered. (TR2:I3)

4. Communal out-of-class tasks that extend and link back into input:

   So this session [on articles] then leads into, ok, then they're divided into groups, and the idea is that each group should have, is given one language area, someone has, there's noun clauses, there's conditionals I think is one of the projects, and there is this series of projects. And the idea is that they should make a worksheet for their fellow teachers . . . come up with something that the others could go through as revision near to exam time, and they post it into communal forums. So part of the idea is, look, there's always stuff to know, and you can't know all of it, but you can help, so part of the message is not just about the grammar, part of the message is, if you were to work as a team you could actually help a little bit with some of this. (TR2:I2)

5. Start (some) sessions with a warmer (but see principle 5 in the following section).
6. Provide answer keys when a right answer exists and it is important for all trainees to know it:

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

I'm doing it [giving out a key] in that session because I think there is a right answer, and I want to be sure they've got it. And, you do get huge variation in the groups, and at least in this way, if you've got people who are really floundering, then, yeah, they've got that bit more information there and then and they can be expert on their questions even if they didn't know what they were before they started the discussion. What, I'm saying it creates an expectation, and perhaps I shouldn't do as much of it as I do, but there are some things where there shouldn't be a key. (TR2:I2)

7. Crossover groups (e.g. the SLA session, discussion of studies from Lightbown and Spada (1993)) – forming groups for one phase of an activity then mixing up the trainees for a subsequent phase so that individuals from different original groups exchange information.
8. Getting trainees to behave as learners (e.g. the session on lexis):

There's a very big tension between the different types of teachers that we have and their different levels of experience. There's an enormous tension between, every time when I get feedback, I will get somebody who loved the fact they had a lot of practical stuff done with them, and somebody who is contemptuous of that, that they were made to behave as a low level learner for 10 minutes . . . but I think it has to be there because there are some people who will not have been in that kind of ELT background. (TR2:I3)

9. Highlighting the relevance of input (to teaching and/or assessment) e.g. highlighting problems with asking learners to guess from context in the lexis session, or explicitly linking it to the Module One exam in the SLA session.
10. Not using technology extensively:

If you've got people in a room, it makes sense to use the people . . . The absence of PowerPoint is actually something of a relief. It's not something I've ever learnt to use. I've sat through so many death by PowerPoint lessons . . . (TR2:I3)

Principles underpinning input sessions

During the interviews, we discussed Victoria's input sessions in order to identify the principles underpinning her pedagogical decisions. Several principles emerged, though it is important to note that Victoria's thinking around these was reflexively fluid rather than dogmatic; she was aware of tensions in her work (e.g. between wanting to minimise lecturing and trainees' demands for direct input) and saw these as a source of motivation for her own ongoing professional development. Training was thus for her: *like a gigantic puzzle. I can't imagine that it will ever stop being a puzzle, and that if it did I think I'd lose interest and stop doing it* (TR2:I3). Key principles which did emerge during our discussions of her input sessions are the following:

1. The principles which underpin good training are the same ones that underpin good teaching:

I think that's one of the things I've already tried to say that you, because you're training you shouldn't walk in there and tell them stuff, you should remember that people, you, it's the same as if you're trying to teach, well anyone English, but you can't walk in and expect people to know things but you need to find out how much they know before you start doing anything with them and you need to vary your focus and pace and you need to make sure they're involved and things like that. (TR2:I1)

2. It is important for trainers to teach to stay in touch with the reality of the classroom: *If a trainer only trains there is a danger of them getting far too up that idealism creek* (TR2:I1).
3. Co-operation and collaboration among trainees enhance learning. She tried to communicate to trainees the idea that *there's always stuff to know . . . if you were to work as a team you could actually help a little bit with some of this* (TR2:I2). Informed by socio-constructivist ideas, she also believed that *people learn together, that they can learn from each other* (TR2:I3).
4. It is important to help trainees understand the rationale for particular pedagogical practices – not just how to use them but why:

I needed to draw their attention more to the meta level of what was going on . . . It's very easy to end up delivering something that's just a string of activities and not, it needs a shape, it needs a reason and it needs something they can take, they can see how it, not only what happened but why it happened and how they can take it through into their practice. (TR2:I1)

5. Only start sessions with a warmer when this is purposefully related to the topic (e.g. the Session 1 picture dictation, linked to work on lexis). She described some of her initial activities as *a hangover from warmers* (TR2:I2) but emphasised the need for these to be purposeful – *we can't be doing with empty warmers* (TR2:I2). She further explained that:

No, I don't use warmers for their own, I don't believe people have to be warmed up. In that sense, you know, why do you have to play a game for five minutes? But if I can find something engaging that works at the beginning of a session, great. (TR2:I3)

6. Deeper learning is more likely when trainees have to articulate their understandings of issues:

I'm going to say it again, it's no use being told stuff, but when you interact with other people you have to articulate stuff for yourself. You have to process it and use it in some way for yourself, because you're interacting with the other people. And that's how you learn, not by having chunks of information delivered to you which are memorised. (TR2:I2)

7. Input sessions should not be spent delivering information trainees can access more directly elsewhere (Victoria was interested in how input sessions might be shaped by current thinking regarding 'flipped classrooms'):

It wouldn't cross my mind to stand in front of them and tell them about articles, they'd go to sleep . . . Yeah. And also, what would be the, engaging is the important word I think, what would be the interest value? Why, I mean, why not, if I'm going to tell them, why not just, but then I suppose why not just look it up instead of coming to see us at all? (TR2:I2)

8. It may sometimes be beneficial to push trainees to engage in learning experiences they initially resist:

So I mean, maybe there is the point at which sometimes you should [push them], because and also I think back, there was a big element when I did my Delta, although it's a very long time ago, there was a big element of [my trainer] getting me to do things that were really a long way outside of my comfort zone that I didn't believe would work. And if he hadn't pushed me into doing them, I would never have understood that they could be done. Or seen any value of it. And having been pushed, you kind of say, ok, right, now I get it. (TR2:I2)

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

9. Trainees should leave input sessions with something new:

   Everyone needs to walk away with something they didn't know. I don't like, when I originally started . . . one of the things I didn't like was there was too much brainstorming and not enough content. And, although it's great for them to share ideas, all of them should walk away with something they didn't, all of them should feel that even, you know, if you're the bright one in the group and you've just, all you've done is share your ideas with other people for an hour and a half, what did you get out of that? (TR2:I2)

10. Input sessions cannot cover all that trainees need to know:

    We're not going to deliver everything they need to know, we're going to just show them what it is they need to know and they need to then work out which of those areas they have to focus on. And this is a big problem for us, that a lot of people arrive thinking, if I sit in this classroom and listen and take notes I will have enough information for the exam. And that's just not true in any way. (TR2:I2)

11. Activities in input sessions should empower and motivate trainees:

    It's a nice, yeah, yeah, yeah, it comes back to that thing about if the person, in the first round [of an activity involving analysis] if there was someone who was totally foxed by the whole thing, they have now [when they receive the key] at least got this little thing that they know. They get to be the person who knows for a little while, you would think would help them, empowering, motivation, and all that kind of thing. (TR2:I2, referring to the use of crossover groups in the session on articles)

12. Advances in online learning mean that input sessions are becoming less important:

    I think this is where it's going. I think the input sessions are going to be less and less important. And it's in a sense right that they should, because we're all living our lives more in this way, this networked way, and it's less about that, and the Delta has to travel with that as well. (TR2:I3)

13. It is important for trainees to understand how input sessions relate to other areas of the course:

    Often at the beginning, they're not able to make the connections terribly well or they're just not bothered about them, I don't know. And if you make the connections for them, you can often sell things that, they need to know why they are doing stuff, you wouldn't want to just sit in a room and do stuff for an hour and not know why you were doing it. So it's one of my pet things is that we need to tell them why and how stuff fits together and what use it will be to them. (TR2:I3)

14. Trainees vary in experience and ability and it is important to cater for this range:

    Delta is, in a way you're dealing with a more mixed ability group than you ever, ever will with any language learners, so you're trying to be all things to all people. So there is some of it, you get, you're trying to keep things very short, because there's some very practical hands-on this is how a lexis thing can be structured, but you've got some people huffing at the front who have been to workshops and I don't know what else and are very much capable of delivering much the same shape of lesson. So you you're trying to keep everything quick, but make sure that you've touched on a structure. (TR2:I3)

15. Cognitive engagement is more important than affective engagement:

    V: I think, certainly the, subconsciously, the cognitive depth things are going on all the time, trying to make sure that there's something in there that's challenging, that keeps them engaged, that keeps them on task.

    S: And does that mean that you're less concerned about the affective?

    V: Yes. [but] I want them to be happy.

    S: So given the choice between making them think deeply and having a good time, you'd go for the think deeply first, and then . . .

    V: Yeah, yeah. And in the general scheme of how our institution runs, that's my position on the, I am the person who makes them work hard.

    S: OK, work hard first, and then entertainment second.

    V: Yeah, as opposed to the entertainer. (TR2:I3)

16. Minimise lecturing:

    Yes, it's [SLA] one of the sessions I started doing very early on when it was effectively a lecture, and I keep trying to change it more and more so that it's less of a lecture, and I think it's, the first time I took it over, it was a session in the notes, rather than one I'd written, and that was a long time ago, maybe 10 years or more. And, but it was a lecture, and I didn't like the fact that it was a lecture, and I've been fiddling with it ever since. (TR2:I3)

17. Communicate ideas without taking a position, publicly:

    I think [my role is] to pass on the information and, it's not that I can't have a position, it's I don't necessarily have to communicate one, and that if I do have a position, there should be, I shouldn't be telling them what to think, I should be telling them things they can think. (TR2:I3)

18. Model good teaching in the way input is delivered:

    And the way it is delivered is that you should be seen to do what you want them to do, so I'm going to go and watch these guys teach for six months, and if they, therefore I should have done what I want to see in their classrooms, and some of them are bright enough to pick that up and some of them you have to point that out. Some of them realise that what you have done with them they can adapt and use in language learning terms, and some of them miss that. (TR2:I3)

19. Getting trainees to work in different groups can vary the group dynamics and distribute the responsibility for peer support across the group:

    *I use crossover groups quite a lot, partly to try and get them all to talk to each other at some point. They tend to come in twos and threes from institutions, and . . . never ever speak to anyone else if you don't do something to intervene with that. Actually, if you can get them to chat to each other a bit in different situations, they will get on. Also you will always get one or two who are much weaker than the others and you need to share them out or they're, if they latch onto one person they can become a real load. Also, you will get one or two people who are a complete pain in the [a***] in any given group, and you ought to share them out as well, because that's life.* (TR2:I2)

Influences on practice and principles

Through the interviews we also explored influences on the development of Victoria's practices and principles as a trainer, with specific reference to input sessions. She did not feel that there were any official constraints on how she delivered input sessions: *There are, the delivery of the course itself, there are very few stipulations. There's only the syllabus document, which is, which is nothing about how or what, it's only, kind of, areas* (TR2:I2). However, several other influences were identified:

1. Own teaching experience:
   Victoria's decisions about how to run training sessions (i.e. process rather than content) were influenced by her experience of what works well for her as a language teacher. As she explained:

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

I think I always just did it the same as I did it with teaching. I'd been teaching at that stage [i.e. when she started training] two or three years and had sort of got better at the idea that you needed to vary the focus and pace, you needed to involve people, you had to feed information in at some point or else they wouldn't go away happy. I think I to a large extent I was just moving. (TR2:I1)

And also:

And to a degree that's partly just me doing what I've always done. That's what I've learned to do from course books. I was thinking this the other day, I'm not, it's not, it's probably no more principled than that, it's stuff that works, it seems to engage people and seems to lead to pennies dropping . . . They engage, and the learner seems to deal with the material and seems to walk away with a better, so it's largely that. (TR2:I3)

2. Observing other trainers, especially her own:

[My trainer] could get people to do things that you just think you can't ask because I'd been teaching in the Water Board and a couple of other government places . . . they were quite stuffy officials and he'd get me to do things and you'd think . . . if he's actually managed to make, I'm not a believer in the playschool approach to teaching but if he's managed to make me run across the classroom in this competitive activity, I can get a 40 year old Water Board official to do it if I just go in there and I did eventually learn that that, as long as you knew it was going to work and you knew how it was going to work and you do it with enough confidence and it did all work and you put the instructions together in the right way so that it worked and actually that they did something then you could make people, so in that sense, yes. (TR2:I1)

3. Ongoing reflection:
   Victoria was engaged in ongoing reflection about her work and this had been a constant source of development for her over the years:

Every time you do anything you want to make it better so you look at what you did constantly in terms of, well if I could just sort that bit out would that make the realisation happen earlier, faster, clearer? And I do that with sessions, I do that with the way I work with the trainees individually and I do that. (TR2:I1)

   She also described her development as a trainer as:

Constant fiddling, constant harking after making things work well. There's always, at the back of your mind, there is always a pile of stuff that you think could be slightly better or a lot better in some way or another and you're prioritising constantly as to what you will have a go at next or what you will change and things work in a sense but you have an idea about how they could work better. So you have to create the time to make them work better and then see if they do or not. (TR2:I1)

4. Her employer:
   Victoria talked positively about the manner in which her employer had got her involved in initiatives which had contributed to her development as a trainer:

Yes, there's another major element that makes me do what I do that I haven't mentioned and that's [my boss], so he does things like walks in, in the summer of 2008 and says, we're going to go fully modular. And I say, we can't, that's mad, no one's trying to do that, everyone's going to try and sell the integrated and then untangle the modules bit by bit. And he says, no, no, we're going to go fully modular because I can sell the modules separately. And I say, don't. And he says, we have to. And then I spend the next six weeks trying to make it work because I know he's going to do it anyway. So there's the being dropped over Niagara Falls element and that also comes from, 'Will you work on an ICELT?', 'What's an ICELT?, yeah ok' eventually when I've, he's explained it's somewhere between a CELTA and a Delta, you say, 'All right that sounds interesting'. And then he doesn't tell you that it's actually going to be with primary school teachers . . . Stuff like that but actually that's really good because actually some of it's been incredibly good for me and working with primary school teachers. (TR2:I1)

5. Socio-constructivist blogs:
   Victoria was interested in some of the ideas related to learning which were being promoted by socio-constructivist thinkers in education generally:

I follow socio-constructivist blogs, I read about how people try to create blended classrooms a lot, and it's something that keeps coming up more and more. And something that, it's also connected perhaps to changes in the way that participants are behaving, and whether, and how far to meet those, and how we can meet those. And the flipped classroom is a very easily thrown about phrase, but I suppose, trying to make sure that we're making the most of the time they are actually with us, and trying to make the sessions more valuable to them. (TR2:I2)

6. Feedback from trainees:
   As already noted, Victoria believed in the value of creating a collaborative and communal learning culture among Delta trainees. Trainees' responses to her efforts, though, were not always positive, and this had led her to question her beliefs about this aspect of her training. This had not, though, at the time of this study led to any major changes in her practices and her response to the more resistant trainees was to *smile sweetly and jolly them into it* (TR2:I2).

Factors which did not seem to impact on her approach to input sessions were:

1. Interaction with other Delta trainers:

This is the problem you see really I've got, I've sort of taken over the practical day to day running of the Delta modules and therefore taken over a lot of the input sessions and so I do what I think is right. And it's one of the reasons I was interested in doing the research project because you occasionally realise that you might actually be an awfully long way from what other people are doing and you ought to know what that is but I don't quite know how you find out what that is. (TR2:I1)

2. The literature on teacher education:
   I asked whether any reading about teacher education had influenced her development as a trainer and her response was: *I suppose I'd have to confess that I almost don't [read]. I'm likely to be looking for things that I think they need to know, not things about training* (TR2:I2). Her reading, therefore, focused more on Delta content, although the blogs on socio-constructivism she followed were noted above.

Case study 2: James (TR3)

James originally did a Bachelor of Education (BEd) in Primary Education before starting his career in Teaching English as a Foreign Language (TEFL) in 1988 and completing his CELTA a year later. He taught in Spain and the Middle East for a number of years before returning to the UK in 2007. He had been a CELTA trainer for 15 years and a Delta trainer for five years.

Features of input sessions

The three input sessions James supplied were: Error correction (Module One, 90 mins), Reading skills (Module One, 90 mins), and First and second language acquisition (Module One, 90 mins). Each of these sessions was taught with a group of six Delta trainees.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

For each session I asked James about the elements of his input sessions that it typified. For the session on error correction the following features were highlighted in our discussion:

- interaction among trainees
- tasks which involve trainees in forms of analysis – e.g. sorting, matching – and discussion, as an alternative to lecturing
- elicitation of contributions from trainees
- recycling ideas from previous input sessions.

James felt that it was *quite fair* to highlight these are typical features of his input work and added: *I don't have a wide range of, what would you call it, sort of matrices in my mind of, as to how sessions should be organised and that, that's, these are, I rely quite heavily on these kind of, on the things you've just described* (TR3:I2).

In reviewing the session on reading skills and strategies James highlighted the following features of his input:

> *I suppose for the whole session then, I suppose the early tasks trying to work out what they know, either by getting them to categorise or brainstorm something, or focus discussion questions where if they don't know any of the terms . . . they can be defined, because they're kind of baseline terms, so that might be something that would happen in the early stages. And I'm trying to get, come to, bring up the concepts there that will be tackled in the session, and that, I'm hoping, I suppose in general then that, and the third, well task two, which is actually the third task, where they look at reading and historical terms, is kind of a little bit of an extra thing really . . . Then the, I suppose, the middle bit, where they do a little more brainstorming on skills and strategies to find out how much they know, or how much they, how much trouble they have in defining these things, and then look at the received knowledge, or at least some of the sources and what they say about them.*

In essence then the repeated pattern here (James agreed) was *brainstorming, so more generally brainstorming followed by some sort of external input* (TR3:I3).

For his third session, which was on language acquisition, James commented on key features of his input work as follows:

> *Well I suppose this one's more heavily theoretical than the previous one, even though the previous one is often called reading theory, it's trying to make a nod towards practice all the way through, and this one is, I think, relatively heavy on theory, I think the concepts, some of them are quite complex, and I hope that we establish early on what their beliefs are and bring up the issues that they have to have beliefs on and kind of force them to make up beliefs, if they haven't made them up already, and I think that sets them up for some, hopefully, in some cases, slightly lecturey type input from me, in some cases, discussion based input, in some cases, task based like in the ordering the natural order hypothesis stuff, and in some cases, gap-filling, and in one case then, in the Krashen case, it's the jigsaw reading thing.* (TR3:I3)

Here, the pattern seemed to be elicitation of beliefs followed by input provided via lecture, discussion and tasks involving processes such as gap-filling and jigsaw reading.

More specific features of James's input sessions were the following:

1. Engaging trainees in interactive tasks and minimising lecturing:
   There was frequent evidence of the use of such tasks in James's materials, although after viewing his video-recorded input session (on error correction) James was critical of how much talking he did:

   > *I'm not enormously enamoured with it, I didn't, I thought when I actually I have to admit when I'd recorded it and I looked through some of the video I thought, well that wasn't fantastic, I wasn't entirely thrilled with it but it's not for me to be thrilled with, it's the study is basically on what happens, not the best of what I can show you. But an average snapshot, so that's one of my sessions . . .I felt that I didn't realise I'd done quite as much talking in it to be honest, and also the timing.* (TR3:I2)

   James also felt that in this particular session he had not been successful in eliciting substantial contributions from the trainees:

   > *I think what I don't do perhaps enough of, this is my feeling maybe, that's why I was a little unhappy when I watched the video, I don't get enough from them, I tend to get a few bits and pieces from them and then start going into lecture mode a little bit too quickly, and I suppose if I asked some more probing questions and some well-chosen questions that could bring out more of the stuff I'm after and their reaction to it and their opinions on it, more effectively.* (TR3:I2)

2. Sequences of task, discussion and input:
   As signalled above, James's input sessions were generally organised around sequences of tasks, discussion and input:

   > *Yes, so in one session there might be two or three of those iterations, although I wouldn't want to make it that boring and predictable. But so it might be sort of, task discussion, feedback with credit to various trainees for bringing up salient points and then a few additional points from myself in mini lecture form, perhaps, yes. Although, I'm loath to think of it in mini lecture form, but I suppose I have to admit that's what it turns out to be sometimes.* (TR3:I3)

3. Trainee discussion of tasks:
   James generally engaged trainees in comparing their responses to tasks prior to any plenary discussion or input:

   > *That's what I was aiming for yeah. And often I'm aiming it, having, not getting the, having them work through some information but, and then perhaps, generally speaking, and this is a bit of a pattern as well, some kind of peer comparison of what they've come up with or something, so I do like them to compare with each other, perhaps in pairs or small groups, before we do whole class feedback.* (TR3:I2)

4. Discussion tasks where trainees share experiences and ideas. The rationale for this practice is outlined in principle 7 in the next section of this article.
5. Follow-up reading, making use of the institution's access to e-journals:
   James was the trainer in this project who more than any other provided trainees with detailed suggestions for further reading, including material of a more academic nature in addition to readings with a practical orientation:

   > *I think yeah, I think possibly, I've been, I've trawled through a fair bit of e-journals as part of my Master's and I suppose yeah that's been helpful. I don't, I find it suitable and I find that it comes back in trainees' background assignments; I don't know if you know what they have to write . . .* (TR3:I2)

Principles underpinning input sessions

During the interviews we discussed James's input sessions in order to identify the principles underpinning his pedagogical decisions. The following key principles emerged:

1. Trainers should model good pedagogical practice:

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

That's [lecturing] not really what we're after, we're after exemplifying, at least this is my take on things. You need to exemplify actually what should happen in the classroom, and lectures are not in what should happen in the classroom, in generally speaking, in my, that's my philosophy at least . . . It's a very close link in CELTA, it's less of a close link in Delta and I'm talking to other Delta tutors, they're a little more understanding about, you know, delivering things lecture format, and I do a little bit of lecturing myself but I try to make it interactive, in terms of there is a discussion task or a problem to solve or something to brainstorm first, and then I'll maybe give a little bit of extra, you know, a little bit of a mini lecture at feedback time if they haven't come up with certain points or if I want to point them towards certain resources that they should read afterwards I think. (TR3:I2)

However, modelling how to teach English (what James called 'exemplification') was not as required on the Delta as the CELTA given the experience of teaching that Delta candidates already possessed.

2. Repeated teacher behaviours define the learning culture:

Every time you go in the classroom, you're establishing or reinforcing elements of the culture of that classroom. You can make the culture over hours and hours of exposure to you and your practice, as you can make them more autonomous or less autonomous, you can make them care more about the other people in the room, or less, so everything that you do should, you should be vaguely conscious of the culture that you're developing or reinforcing. So, for example, crediting students with coming up with the right answer is positive, taking the credit yourself, so saying, yes good, well done Ahmed, so let's have a look at what Ahmed said, and then it's not real until it comes from the teacher, that's perhaps taking away autonomy from learners. (TR3:I3)

James explained that he hoped that reinforcing positive teaching behaviours through his own practices was something he did as a trainer.

3. Delta trainers need to ensure there is sufficient theoretical substance to sessions:

James was one of the trainers who often referred in our discussions to the need for theory on the Delta. He felt that: *There's a definite need for a certain level of input, more meat on the bone basically, more substance, so more theoretical substance and signposting towards more, towards places where trainees can get more theoretical substance.* (TR3:I2)

4. Trainers need to be able to walk the walk not just talk the talk:

There's a certain professional pride in I suppose in you can't possibly tell people what to do, you don't tell what to do, you show people what to do. I think in, even in the classroom you are always a little bit wary of telling people what to do and not actually putting up yourself and actually doing it yourself . . . If you constantly say 'you didn't check your instructions', well then you have to exemplify a few ways to do that and do them regularly, 'you didn't check concept on that grammar you presented', well you have to exemplify how that can be done. (TR3:I2)

5. Trainees should interact critically with the input they receive:
The need to develop criticality among the trainers was a concern highlighted by several trainers in this study. James's views on this issue were:

Well I don't necessarily ever in, well very rarely in any case in input, want to get to a stage where it feels as if I'm saying, it's all wrapped up, so I've just told you how it is, so that might be with the second language acquisition theory, which definitely, certainly isn't all wrapped up . . . so I want to leave it open-ended and I do use hedging language when I talk about, you know, well it may be that and we use, many people believe that it may be the case that, and I think the job of the theory sections of Delta is for them to get a good list of these 'mays' in their heads, so that they can carry them over into the classroom or into their reading and test them out themselves. (TR3:I3)

6. It is important to ensure trainees are prepared for the assessments they do on the Delta:
James referred to assessment at several points as an influence on how he conducted input sessions, and in our final interview he noted that: *As you've probably seen, I'm very driven by the assessment* (TR3:I3).

7. Trainees benefit from opportunities to share experiences and viewpoints:
As already noted, a recurrent feature of James's work was the use of discussion tasks where trainees compared experiences and ideas. I asked him about the rationale for this practice and this related mainly to the benefits that trainees derive from learning about different perspectives on teaching English:

They come from different backgrounds, some of them work in the FE sector, some of them work here with us in the HE sector, some have been abroad and done quite a bit of work in private language schools and some have never done that, and they've all taught, there's a broad variety of, between them of different L1 speakers and there's one non-native Delta candidate there as well, so they have a lot to tell each other. They have a variety of approaches to things, they are quite different in the classroom, I've seen now all of them teaching and they do work quite differently in the classroom . . . so, because often they'll hear something from a peer that then I might say later on, or if it's been brought up quite comprehensively across the two or three tables then I won't have to say it really, they'll say it to me maybe. (TR3:I2)

8. Eliciting trainees' beliefs makes them more receptive to the subsequent input:
As noted above, during input sessions James typically went through cycles of eliciting trainees' beliefs prior to providing input. I asked him for the rationale behind this approach:

Well, they're not the only way I do things, obviously, but they are, I suppose, one of my favourite ways, or one of the ways I've come lazily to rely on, perhaps might be more accurate, but the idea is, that you've got your Delta trainees who have, at the very least, two years, a pre-service course under their belts in almost all cases, and two years of active teaching, or more, in many cases. So they should have a lot to bring to the party, they should have quite a bit of, if they've had any reflective time on what they do, they should have already some ideas about what they think about, oh, I don't know, reading strategies or language acquisition or error correction, or whatever it might be. And so that may well link quite neatly to or perhaps even preview some of the more theoretical input that I'm trying to get across to them, so I'm wanting them to become more receptive to this input, or in some cases, bring up the concepts themselves. (TR3:I3)

9. Acknowledging what trainees already know is an important way of enhancing their confidence:

I think it's important, because people come on the course with a fair bit of expertise, practical mostly, built up over a few years of teaching and you don't want to, I don't, one thing you really do want to avoid is creating any kind of mystique or shaking their confidence by saying, well look at all the stuff you don't know, you know, that's really not the object here because they're all, if they've reached this stage, unless we've really misjudged things, they're all very capable of doing well on the Delta course, you know, so. (TR3:I3)

10. Trainees need to be directed towards certain knowledge – only eliciting what they currently know will not suffice.
Although James gave plenty of weight to discussions

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

where trainees shared their current views on an issue, he felt it was important to go beyond such discussions in order to ensure that trainees were familiar with the knowledge they needed to know:

> *Certain points of the compass have to be touched on, there might be a list that may not be a finite list of terminology but there's a list of terminology that they perhaps have ideas for in their heads but they need to attach those to certain terms because they will come up in Module One and they will be expected to come up, perhaps in their background assignments as well, for example, just as an example, and also you do, to some extent it's not purely metalinguistic, or it's not just terms, it's also they may have perhaps a slightly limited view of a certain area, like say, the reading skill, and you may want to try and expand that a little bit, so it's not just enough for them to come up with and to defend their own concept of the whole area, you need to challenge that a little bit, I think sometimes.* (TR3:I3)

Influences on practice and principles

Through the interviews we also explored influences on the development of James's practices and principles as a trainer, with specific reference to input sessions. He did not feel that there were any official constraints (i.e. administrative constraints or constraints imposed by Cambridge English) on how he delivered input sessions, but several other influences were identified.

1. Own teaching:
   Once again, his own training and experience as a teacher was an influence on his practices and thinking as a trainer:

   > *No, I think I do what I've been trained to do and I've been trained to teach English, and I think I do teacher training largely in the way that I've been trained to teach. And it's something that I think that's fairly defensible to be absolutely honest with you.* (TR3:I2)

2. Training up (CELTA/Delta):
   James described ways in which being trained up as a CELTA and Delta tutor had influenced his own practices as a trainer. He agreed that much of the training up was *more about assessments and standards and procedures and feedback to candidates, to trainees rather and making sure that they are aware of how they're doing and aware of what they need to achieve* (TR3:I1) but he did also recall having to observe other trainers giving input and to describe what they did using an observation form in the training-up handbooks.

   > *The trainer trainee handbook has a sheet that you fill out while you observe input and towards the end of it you tick, it's a little tick box thing and that's when I discovered what loop input was, I, that you have to check how is input delivered? And loop input was one of those things that was in there.* (TR3:I1).

   He also recalled how exemplification (i.e. where the trainer taught a lesson and the trainees pretended they were learners) was a technique he had seen trainers in both CELTA and Delta contexts use:

   > *It was modelled in our, in my work history and my take on the whole profession, both in my own CELTA course and in my own Delta course and then in observing trainers when I trained up to do CELTA and ditto when I trained up to do Delta, this has always been present.* (TR3:I2)

3. Experience as a CELTA trainer:
   James was a CELTA trainer before starting to work on the Delta and he felt that his earlier experiences carried over when he became a Delta trainer: *Quite a lot I have to admit, quite a lot, the way sessions were initially organised etc.* (TR3:I1). In fact he talked about establishing a pattern for running input sessions quite early in his career and which had continued to be influential:

   > *J: I quickly enough developed a way of doing things and I feel it's quite difficult to shake out of that sometimes, as a procedural way to run an input session. I think it works, but I think it would be nice to inject a little variety in the way I approach things sometimes.*
   >
   > *S: So has that carried through to your work today you mean?*
   >
   > *J: Yes I think so, I think I've just got better at doing it the same way and in some respects it might be better if I try to do it a different way sometimes . . . if I thought it didn't work I would obviously go back to square one, but I think it has its advantages. I think it's never good to do too much of the same thing, which is why I would like to try something a little bit different sometimes. (TR3:I1)*

4. Observing new trainers:

   > *The trainers I've managed to watch, they are initially the people I shadowed when I was training up and then subsequently I'm quite lucky here I guess because we have quite a few trainers now, we've trained up quite a few people and even if people train up at a centre, even if they were all to just observe the same CELTA tutor as part of their training, they end up as different tutors with different ways of doing things. And I do as, because I manage the teacher training side of things here and I do get to see them deliver input and we have some quite talented people here and often they do things a little differently. The overall aim is the same and often it will be the same session they're delivering, but they will do things very slightly differently and I will think, well why don't I do it like that any more? Or should I do more of that?* (TR3:I1)

5. Talking to trainers:

   > *Yeah, talking with other trainers as well. Yeah talking about lessons you've seen, teachers you've seen, you wouldn't see them, you wouldn't both see the same lesson, but you would both, if two tutors ran a course and there has to be two tutors on a course, you would both see the same trainee. So talking about trainees and what they do and don't do and what our reaction to that is quite a lot and that changes practice I think as well.* (TR3:I1)

6. Delivering other trainers' sessions:
   James did on occasion have to base input sessions on materials developed by another trainer and he found the process of engaging with and adapting these materials a developmental one: *Delivering sessions that somebody else has designed, or at least getting somebody else's sessions designed from a previous year and adapting and changing it and thinking, oh I'd like to do this and that and the other.* (TR3:I1)

7. Reflection:
   James also highlighted ongoing reflection as another influence on his development as a Delta trainer:

   > *And also basically reflective practice I suppose if that's the way, what you'd like to call it, basically because we've been lucky enough to run courses for the last three, for three and bit years now and you get to do things again and you don't do them the same really, unless you thought, well it wasn't so bad. But you will focus on at least a fair chunk of the course that you want to adapt, but never throwing it out and starting from scratch, I wouldn't have the heart basically to start developing a new session from scratch it would, it's just incredibly time consuming.* (TR3:I1)

   This sense of regularly reflecting on the effectiveness of input sessions emerged strongly in the next quotation too:

   > *You reflect after the session, I don't, I should I'm sure, keep a journal of some sort, but I don't reflect in writing after each session. But you do reflect afterwards*

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

and you think well that, maybe these sections didn't take as long as I thought they would, took too long, proved too difficult, you look back on something and you think, I don't know how I didn't see that but there was just too much discussion and not enough input or whatever and so that's one of the ways that I would make decisions on changing a session. (TR3:I2)

8. Assessment requirements:
   James noted above that ensuring trainees were ready for assessment was an important concern for him. Assessment and the syllabus were in fact broader influences on how he had come to think about input sessions. For example, he often used tasks that mirrored those the trainees would have to do in the Module One exam:

   We, I think the requirement that we, at least what we consider to be the most important requirements are, I mean it's the backwash effect isn't it? You know what they have to achieve, so you want to make sure that your input sessions support them in achieving that . . . Yes the syllabus is a powerful influence, the assessments are the most, for me personally the most powerful influence. So you would have to do a, quite a lot on course design for example because Module Three is on course design yeah. (TR3:I1)

9. Trainee performance:
   The general level of trainee performance during the Delta was another influence on decisions regarding input sessions cited by James:

   Looking at what they produce later in the course, or right through the course and they produce a lot of text for Module Two background assignments, they produce lesson plans, they actually teach lessons and you watch them and obviously then they do bits, other bits and pieces as well and we do . . . we do mock questions now and then for the exam and if they're producing stuff that indicates a lack somewhere, well you might try and trace it back to a session. (TR3:I2)

   Finally, James highlighted a lack of alternatives as a reason why his approach to input sessions relied, he felt, on a small set of long-established routines. He had not found online options to be of particular value in developing his input sessions and was looking for more practical face-to-face strategies he could use:

   I've done a little bit of work on e-moderation and e-learning recently but it doesn't entirely convince me. It seems to me like something that doesn't add value but helps overcome problems of distance but it doesn't really add value. I would like, I'd like to find, I'd like to discover some new ways of doing things face to face, I'd like to experiment, but I'm at a little bit of an impasse, I can't think of any in particular . . . what I'm after here obviously is very practical, I can conceptualise what the content of an input session needs to be. But I'm after maybe a couple of alternative ways to actually, not deliver that's the wrong word, but to get them through it, to help them explore it. (TR3:I2)

General findings

The above cases illustrate within the space available for this report the kinds of insights that this study has generated about the nature and origins of Delta trainers' practices and principles. Drawing on the above cases as well as the other five that were studied in this project, the following general conclusions emerge:

1. Despite working in very different contexts, the trainers held a common core of beliefs about learning generally and L2 teacher learning specifically. Shared principles which informed their work were:
   - trainee-centred activities during input sessions promote learning more effectively than lectures
   - Delta trainees have much valuable knowledge and experience which should provide the starting point of input sessions
   - even with more theoretical sessions it is important to help trainees consider how input is relevant to classroom practice
   - it is important to prepare trainees adequately for the assessments they will do on the Delta
   - trainers should be models of good practice in teaching
   - opportunities for trainees to share and compare experience and ideas facilitate teacher learning
   - at in-service level trainees should not be given dogmatic advice but supported in developing a critical awareness of different ideas and pedagogical strategies
   - trainees need to understand not just how to teach but the rationale for their pedagogical choices; they also need to be able to articulate this rationale.
2. The manner in which the above principles were transformed into training pedagogy varied across the trainers, though again some common general patterns were discernible. For example, there was a preference for cycles of activity during input sessions which had this structure:
   - exploring trainees' current practices, knowledge, or beliefs
   - one or more worksheet-based interactive analytical, discovery or problem-solving tasks (e.g. matching, sorting, gap-filling)
   - trainee discussion of responses to the task (often in new 'crossover' groups)
   - plenary discussion
   - mini lecture (often considering the applications of theory to the classroom).

   There were, though, several areas of practice where the trainers varied. Examples of such areas are:
   - how much direct input (lecturing) was appropriate during input sessions
   - whether trainees should be forced to change who they work with or left free to decide on their own groupings
   - how widely technology (e.g. PowerPoint, video, online discussions) should be used during and after/before input sessions
   - how much further reading to assign or to expect of trainees after input sessions
   - whether trainees should be given 'answer keys' to the tasks completed during input sessions
   - whether warmers should be used at the start of input sessions
   - how to balance attention to assessment and to professional development more generally
   - how to bring input sessions to a close
   - how to evaluate the effectiveness of input sessions.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Trainers were generally consistent in aligning their practices and their principles; there was only one case where a trainer felt his video-recorded input session did not sufficiently reflect his principle of minimising lecturing. There were, though, variations in the depth with which trainers were able to articulate the principles behind their work. For example, although interactive tasks in pairs and groups were a central feature in the work of every trainer in this project, justifications for it varied from references to socio-constructivist theory, to the more immediate value to trainees of hearing what others think about a particular issue, as well as to the notion that learning is deeper when trainees have to articulate their views in a coherent manner to others. Some trainers were challenged by 'simple' questions such as 'why do you ask the trainees to work in groups?' The source of the challenge was not so much the question as the fact that the trainers had never been asked to articulate their rationale for such an embedded and taken-for-granted feature of their work. Overall, the analysis of trainers' practices and principles suggests that the relationship between the two is not a linear one, and that similar practices during input sessions may be underpinned by different principles.

3. The third research question addressed here related to the factors that had shaped Delta trainers' current practices and principles during input sessions. A wide range of influences emerged, though two recurrent themes in the work of the trainers were their own experience as teachers and their own Delta training. Regarding the former, the trainers often talked about training as an extension of teaching and about the manner in which becoming a trainer meant transporting from the language classroom to the training room those pedagogical practices which they had found effective in teaching English. A second powerful and lasting influence on these trainers was their experiences of doing the Delta and the manner in which their trainers worked with them. In fact, one explanation for the common pedagogies and philosophies which this study has highlighted among this globally dispersed group of trainers is the fact that many of them did their Delta at the same institution. In this sense, the Delta training highlighted in this study has a strong sense of tradition attached to it. This may tend to promote conservatism in the way input sessions are run as trainers perpetuate, in their own work and in those of new trainers they work with, practices and principles established long ago.
4. In addition to the two key influences on trainers noted above, the following factors also impacted to varying degrees on what they do and how they think in relation to input sessions:
   - ongoing reflection
   - socio-constructivist blogs
   - line managers
   - feedback from trainees
   - previous training experience
   - observing new trainers
   - sharing training materials with colleagues
   - talking to colleagues
   - syllabus and assessment requirements
   - L2 learning experience
   - reading.

   It should be noted though that talking to colleagues, where this was mentioned as an influence on trainers' practices, was an internal process; dialogue with Delta trainers from other institutions was an element which was lacking in the experience of the trainers in this study, other than where this related to assessment standardisation work. The chance to talk about the pedagogy of input sessions was thus a professional development opportunity that these trainers had not had. Reading, too, did not emerge as a significant influence on how trainers approached their input sessions; their reading related more to the content of the sessions than to their own pedagogy.

Conclusion

Drawing on video and documentary evidence of Delta trainers' input sessions and on their own oral accounts of the principles behind these, this project provides valuable descriptive and interpretive insights which will contribute to existing research on language teacher education in general and in particular to extending the limited empirical accounts that exist of what trainers do and why. The case studies this study has generated and the practical insights that are derived from them can also provide the basis of professional development support for Delta trainers. In fact, one key finding here relates to the lack of such support which these trainers have experienced; while, as this study shows, they have all drawn on other sources to further their own professional growth, the absence of opportunities to learn about what their peers do and to talk with them about the pedagogy of input sessions is clearly an issue that merits consideration. Finally, it is also worth acknowledging here that the conversations the trainers had with me were for them an opportunity to reflect on their work and to talk to an outsider about it in a way they had never previously had the opportunity to do. There were various examples during the study of how these conversations created awareness in the trainers or stimulated them to think more profoundly about areas of their work. I have already mentioned above the trainer who realised from the video of one of his sessions that he lectured more than he thought; another trainer became aware of the potentially repetitive structure she followed for all the language awareness work she did, while a further example was of the trainer who had never really thought about how he brings input sessions to a close. One purpose of the study was to provide trainers with a developmental opportunity and I think that in this respect the project was successful.

Acknowledgements

I am grateful to Cambridge English Language Assessment for funding this project and to the participating trainers for their interest in and commitment to the study.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

References

Bates, T, Swennen, A and Jones, K (Eds) (2011) *The Professional Development of Teacher Educators*, London: Routledge.

Borg, S (2006) *Teacher Cognition and Language Education: Research and Practice*, London: Continuum.

Borg, S (2011) Language teacher education, in Simpson, J (Ed), *The Routledge Handbook of Applied Linguistics*, London: Routledge, 215–228.

Cambridge ESOL (2009) *Delta Module Two: Training and Induction Handbook*, Cambridge: Cambridge ESOL.

Johnson, K E and Golombek, P R (Eds) (2011) *Research on Second Language Teacher Education*, London: Routledge.

Koster, B, Brekelmans, M, Korthagen, F and Wubbels, T (2005) Quality requirements for teacher educators, *Teaching and Teacher Education* 21 (2), 157–176.

Lightbown, P M and Spada, N (1993) *How Languages are Learned*, Oxford: Oxford University Press.

Washback effect of Cambridge English: Key for Schools in São Paulo city, Brazil: A case study

MATILDE VIRGÍNIA RICARDI SCARAMUCCI UNIVERSITY OF CAMPINAS, BRAZIL
ELIANA KOBAYASHI UNIVERSITY OF CAMPINAS, BRAZIL

Introduction

English proficiency exams, especially those produced by Cambridge English Language Assessment, have been increasingly popular in Brazil, administered for job selection, university requirements and to enhance a curriculum vitae. Consequently, helping students to pass such exams has become a matter of concern for English schools, regular schools, teachers and other professionals in the field. Materials, lessons, strategies, and techniques thought to help both teachers and students are used intensely. On the other hand, collecting evidence on how such exams influence teaching has received little or no attention in this country.

Many Cambridge English exams are popular in Brazil but we chose to analyse whether *Cambridge English: Key for Schools* (also known as the *Key English Test (KET) for Schools)* has influenced teaching in a private school in São Paulo city, especially because of its recent development as a version for younger candidates of *Cambridge English: Key* (also known as *Key English Test (KET)*). Although it is believed that teachers' activities in class have been influenced by exams, there is no evidence to confirm this claim.

This study investigates the washback effect based on the hypotheses presented by Alderson and Wall (1993) which suggest: 1) a test will influence teaching; 2) a test will influence what teachers teach; 3) a test will influence how teachers teach; 4) a test will influence the rate and sequence of teaching; 5) a test will influence degree and depth of teaching; 6) a test will influence attitudes to the content, method, etc. of teaching; 7) tests that have important consequences will have washback and conversely; and 8) tests that do not have important consequences will have no washback.

In order to better specify the nature of such influences they will be analysed according to the washback dimensions presented by Watanabe (2004). The first one refers to the specificity (general–specific) of the effect. If the test produces an effect that is not specific to it but expected to be produced by any other exam, it can be considered general. On the other hand if the effect is due to some particular aspect of a test, the washback is specific. The second dimension conceptualised by the author is related to the intensity of the effect, which can be classified along the strong–weak continuum. The former means that what happens in a class is determined totally by the test while the latter means that only some aspects of classroom activities are affected. Concerning the third dimension, length, Watanabe discusses how long the influence can last. It is short if after the exam it disappears and long if it continues even after the test. Although every exam is expected to have certain effects, which are usually called intended effects, exams may also produce unintended effects or washback. This would be referred to as the washback intentionality dimension. Finally the value dimension is related to whether the washback is considered positive or negative. However, the author emphasises that such a judgement depends on who is evaluating the effect.

Methodology

Research questions

This research aims at investigating if and how *Cambridge English: Key for Schools* influences the practice of one Brazilian teacher of English in a private bilingual (Portuguese–Spanish) school which offers both elementary and secondary education, answering the questions:

1. What understanding of the *Cambridge English: Key for Schools* exam does the Brazilian teacher of English have?
2. How does the teacher interpret the exam construct?
3. Does the exam have any influence on the English taught as a subject at school? If so, what is the nature of this washback?

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Context and participants

School profile

This study focused on an 8th grade class of a private bilingual (Portuguese–Spanish) school which offers elementary and secondary education, located in an upper-class neighbourhood in São Paulo city. This school was among those identified as having a high number of *Cambridge English: Key for Schools* candidates in the city. According to the English co-ordinator the exam is usually taken by students who are around 12 years old and are in the 8th grade. There are about 1,500 students attending the school on a full-time schedule – classes take place in the morning and in the afternoon.

Student profile

Although most of the students are from upper-class families who can afford high tuition fees, there are exceptions, such as the children of school staff who are offered full scholarship. The majority of students usually travel abroad and have regular contact with the English language, for example by playing video games, surfing the internet and watching movies.

English was introduced to the elementary levels in 2002. All the lessons are delivered in English and the four skills – reading, writing, listening and speaking – are considered equally important.

Nowadays the three 50-minute weekly lessons in 8th grade are divided into three parts: the grammar/vocabulary lesson is given to the whole class; and for the listening/speaking and the reading/writing lessons the class is split into two smaller groups. It is believed that students may better develop their language skills in smaller groups as teachers can give them individual attention more easily. The aim of such division is only to reduce the number of students per class so other factors, such as level of proficiency, are not taken into account. Consequently both groups receive the same type of instruction. The grammar content follows what is proposed in the coursebook but the teacher can complement it when needed. There is an average of 35 students in an 8th grade class.

Teacher profile

All the English teachers of the school have the *Cambridge English: Proficiency* (also known as *Certificate of Proficiency in English (CPE)*) certificate and the 8th grade teachers have been in the school for more than 15 years. They do not receive any sort of training on Cambridge English exams but as there are many teachers in the school who are also Cambridge English Speaking Examiners, they end up giving advice on the oral part of the exam to their colleagues. Weekly and monthly meetings are held so teachers can discuss teaching objectives and maintain similar behaviours in class which are in line with the school's educational beliefs, such as encouraging students to work in a collaborative instead of competitive way and fostering values like respect among each other.

English programme

The English course programme is based on Parâmetros Curriculares Nacionais (National Curriculum Parameters), the Common European Framework of Reference (CEFR) (Council of Europe 2001) and on the school philosophy, which aims to enable students to understand, act and transform their society. According to this document, English language teaching should not be limited to the teaching of language forms, isolated from context with its social, cultural, political and ideological values but based on a social and interactional view of language and learning.

The general teaching objective is divided into five levels: 1) pre-school; 2) elementary education I – 1st Cycle (2nd and 3rd grades) and 2nd Cycle (4th and 5th grades); 3) elementary education II – 1st Cycle (6th and 7th grades) and 2nd Cycle (8th and 9th grades); 4) secondary education (1st and 2nd grades) and 5) secondary education (3rd grade).

Among the English teaching objectives of the elementary education II (1st and 2nd Cycle) are: 1) encourage and develop sociability, co-operation, responsibility, initiative, respect and organisation; 2) raise students' awareness of their engagement in the teaching and learning process; 3) broaden students' minds by exposing them to the habits and culture of the people where the foreign language is spoken; 4) develop students' competence to use the language as a way to communicate in oral and written form, and their knowledge of the language as a system to use it in dialogues, descriptions and narratives; 5) develop reading comprehension to understand specific and general ideas in texts according to different language levels; 6) develop skills to infer and establish relationships, justify and classify, analyse and summarise; and 7) to relate visual comprehension to oral and/or written comprehension.

The course objectives for 8th grade students are: 1) understand, speak, read and write about past situations; 2) describe different lifestyles considering cultural variety; 3) talk about household chores; 4) ask and give directions; 5) ask for permission; 6) make invitations; and 7) talk about personality.

To achieve these objectives, the programme content is divided into conceptual content, procedural content and attitudinal content. The conceptual content, on the other hand, is divided into lexical and grammatical items. The former consists of lexical items related to city places, household chores and personality while the latter consists of simple past, progressive past, short and long comparative and superlative adjectives, *must/mustn't*, *have to/don't have to*, *some/any/no/every*, and *be going to*.

The procedural content comprises: 1) revision and use of verbal tenses in contextualised activities; 2) functions, vocabulary and grammar of dialogues and written texts; 3) reading and listening comprehension related to topics studied in class; 4) preparation and presentation of oral and written texts; 5) dictionary use to understand new vocabulary; 6) guided written tasks; 7) use of functions, vocabulary and grammar content; 8) reading comprehension of readers using reading strategies; and 9) participation in immersion activities.

The attitudinal content must cover: 1) grammar as a tool to allow good communication; 2) vocabulary acquisition; 3) awareness of the importance of communicating in English; 4) interest and pleasure in communicating in the language; 5) interest and pleasure in taking part in drama situations; 6) participation in individual, pair and group work; 7) value

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

of the sociolinguistic behaviour which contributes to group relations; 8) raising students' awareness of the importance of understanding the whole idea of oral and written texts; 9) interest and curiosity of understanding new ideas expressed in written texts avoiding pre-conceived ideas which can interfere with understanding; and 10) interest in reading written texts in an independent way, aiming at getting information to broaden knowledge and also reflecting in order to produce meaning.

For the 6th to 9th grades the course programme document defines the teaching approach as communicative–functional and emotional–affective which allows students to develop the competence to use vocabulary in an active way as well as use the language structure to express ideas and wishes, so content should be taught as a tool to reach the socio-interactional view of language.

For the lessons, activities and tasks, teachers can use different strategies, which they can change according to the group needs and the teaching/learning situation. Comprehension strategies such as skimming, scanning, main points, prediction, cognate words, previous and world knowledge are recommended.

Assessment is seen as a continuous process focused on qualitative aspects. It should: 1) be contextualised and focused on reading, listening, speaking and writing; 2) develop grammatical, strategic, sociolinguistic and discourse competence; 3) distinguish formal and informal language and identify in which contexts they are used; 4) relate texts and contexts, taking into consideration interlocutors, time, place, and technologies; and 5) see texts in a coherent and cohesive way where expressions and vocabulary are used according to social-cultural aspects and communicative purposes.

The Cambridge English: Key for Schools exam

According to the exam handbook (Cambridge ESOL 2012), *Cambridge English: Key for Schools* is aligned with Level A2 of the CEFR (Council of Europe 2001). It is the same as the *Cambridge English: Key* exam other than it is aimed at younger candidates in terms of content and topics.

The test has three papers, each one consisting of a different number of parts: Paper 1 (Reading and Writing) with nine parts; Paper 2 (Listening) with four parts; and Paper 3 (Speaking) with two parts. Each part has different task types, focuses and number of questions.

In Paper 1 (Reading and Writing) the candidates have to read texts of different genres such as notices, dialogues and short texts in order to carry out tasks which involve matching, multiple-choice cloze, open cloze and word completion. The test assesses the candidate's understanding of general and specific information and knowledge of grammar structure and vocabulary in contextualised situations. Regarding writing, candidates must complete an information transfer task by completing a form or notes with a short answer. They must also write a short note, email or postcard in response to a written prompt which includes three content messages, all of which must be covered in the response.

Paper 2 (Listening) assesses candidates' ability to identify simple factual information in dialogues and monologues of different lengths by requiring them to answer multiple-choice and gap-filling questions.

In Paper 3 (Speaking) candidates take the test in pairs. During the first part of the test, candidates interact with the examiner by answering questions on familiar topics. For the second part of the test, candidates interact with each other by asking and answering questions about daily or social life based on prompts.

This analysis reveals some aspects which are salient and more likely to influence teaching. Format, for instance, is one of them: all papers have specific formats which do not change. Another aspect is the genres involved in Paper 1, which are defined as signs or notices, newspaper and magazine articles in the Reading tasks and note, email or postcard in the Writing part.

The handbook also presents the exam specifications, which include language purposes, an inventory of functions, notions, grammar, lexis and topics, offering teachers information to prepare students for the exam (Cambridge ESOL 2012).

Further analysis of the papers shows that in many parts, assessment is based on the integration of skills: 1) in the Writing part, candidates have to be able to understand written texts (i.e. read) to complete a form and a short text; 2) the Speaking paper requires reading, as candidates must read a prompt card which presents written and illustrated information before interacting with another candidate, and also listening as candidates must respond to the examiner and the other candidate. In assessing speaking skills, the tasks involve being able to interact with the examiner (a more capable interlocutor) and also ask and answer questions with the other candidate who is likely to have a similar proficiency level.

Moreover, *Cambridge English: Key for Schools*, as with most international language proficiency exams, requires candidates to transfer their answers to an answer sheet, a procedure which is not usual in school achievement tests except in mock examinations.

Coursebook format

The coursebook for the 8th grade is organised into eight units and its content comprises vocabulary, grammar, communication and skills. The topics are related to places around town, housework, personality, the weather, exciting experiences, books, crime and human achievement.

All units are organised in a similar way. First, a reading text with comprehension questions is presented, followed by vocabulary and word definition activities. Grammar comes next, in which rule explanations are followed by exercises with varied formats such as gap-fill, sentence rewriting and matching. The units always close with activities in reading, listening, speaking and writing on the same topic of the unit. They also start with a reading text followed by comprehension question-answer tasks. Then comes a listening activity with different formats. Speaking tasks always involve pairwork in which there are occasional reports to the class. The writing tasks focus on different genres such as emails, chart completion, book review and descriptions.

Data collection

The data for the study comes from classroom observations, analysis of the *Cambridge English: Key for Schools Handbook for*

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Teachers (Cambridge ESOL 2012), the course programme and materials used in class, as well as a semi-structured interview with the teacher.

The data collection was divided into two phases. In the first phase, the 8th grade English lessons were observed and just when this phase was over the teacher who taught those lessons was interviewed. Such division was necessary to avoid influencing her performance in class.

Classroom observations

A total of 14 lessons were observed and audio-recorded over eight consecutive weeks. Seven observations were done of the whole class and seven of the smaller divided classes. An average of two lessons was observed weekly.

The classroom observation tool was designed for this research based on the most salient aspects of the exam handbook. The categories were: skills (genres, patterns of interaction), task types, grammar content and functions, notions and communicative tasks. The observations were done by the researchers and the collected data was triangulated with other data (i.e. interviews and document review).

Interview with the teacher

The observed lessons and the interview were conducted with an English teacher who was also the English subject co-ordinator. She graduated in Lettres from a traditional private university and has worked for the school since 1994.

The semi-structured interview, which was audio-recorded and lasted 1 hour, aimed to focus on four main topics: 1) identify her views of the *Cambridge English: Key for Schools* exam, as well as her views of language and proficiency; 2) find out how the lesson plans and the students' assessments are conceived; 3) check if the students are encouraged to apply for the exam and if this is the case, how; and 4) construct student and school profiles. Thus the script was divided into four parts: 1) the Cambridge English exams; 2) the classes; 3) the assessment; and 4) the school, the students and the teachers.

Both researchers participated in this collection. The data was analysed based on objectives set for the interview and afterwards it was triangulated with other data collected (i.e. observations and document review).

Findings and discussion

This section presents the main findings from the data collected from each instrument and a triangulated analysis which allowed us to answer the research questions.

Classroom observations

The observed lessons showed that the teacher uses English at all times during class lessons and so do the students when talking back to her. Apart from the students' coursebook, chosen by the school, the teacher uses a range of complementary materials from YouTube videos to English CD-ROMs.

Most of the time, the skills were developed in an integrated way. In two of the observed lessons students were asked to present a summary of a short story, orally and individually. In another lesson, the teacher integrated speaking, listening and reading. Using a video from YouTube, she started a group discussion which was followed by a reading comprehension activity.

It was observed that regardless of the lesson objective, the oral interaction pattern was mostly teacher–students apart from one lesson, when students interacted in pairs following a reading text with questions and answers.

The listening and writing skills were also taught in an integrated way. One of the materials used was a CD-ROM collection which comprised five episodes of about 6 minutes each. Students had to write answers to written questions after watching the videos. The instructions and all the questions were written in English and students were required to answer in English. In addition, an American TV series called *Smallville* was also used to promote writing narratives.

In the observed lessons, the *Cambridge English: Key for Schools* exam was never referred to by the teacher or students.

The Cambridge English exams

Cambridge English exams were first administered in the school more than 10 years ago. The school decided to start administering the exams to meet some parents' expectations as they themselves hold Cambridge English certificates. All the students are invited to take the *Cambridge English: Key for Schools* exam but usually it is up to the families to decide if the children should take it. The school advises the family if the chances of passing the exam are low. The fee to take the exam is paid by the family separately from the school fees.

When interviewing the teacher, we also learned that in June all 8th grade students started doing the *Cambridge English: Key for Schools* Reading, Writing and Listening mock exams as part of their achievement tests. This allows teachers to determine which students are likely to pass the exam and better make recommendations. In the teacher's opinion, the mock exams help familiarise students with the exam format and procedures such as transferring answers to the answer sheet. The mock exams are taken from the digital version of the coursebook. Once the students are registered for the exam, they start attending extra classes to be prepared for the Speaking paper.

The students identified as ready for the exam can take it earlier in the year while those identified as still in need of more preparation are advised to take it later in the school year.

Most of the students in the 8th grade also take extra English lessons. Although co-ordinated by the school, the lessons are offered at the premises by outsourced professionals and it is necessary to pay extra tuition fees. Currently there are 300 students attending extra classes.

The teacher demonstrated her strong understanding of the expected levels of proficiency needed for *Cambridge English: Key for Schools* and other Cambridge English exams. Reflecting on some past candidates she says: *For example, this student . . .* (student's name) *. . . she is now on 8th grade and she has already passed Cambridge English: Key for Schools . . . like this* (student's name) *. . . now this for example was already on 9th grade . . . but he was not mature enough for PET [Preliminary English Test, now known as Cambridge English: Preliminary] for*

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

example . . . it depends on the student. We observe this because what's our intention? It's . . . they go to the exam with confidence, self-esteem, motivation.

The teacher says she noticed a change when the *for Schools* version of the exam was introduced. When the students took the *Cambridge English: Key* version, they had to be prepared to manage situations and expressions they were not familiar with. She recognises the *Cambridge English: Key for Schools* exams are more suited to the adolescent world, which has made preparation more natural since they are related to the adolescents' reality. *Before, we had to prepare our adolescents with comprehension, precocious expressions for their age to take an adult exam. So we had to get them matured for this exam and now not anymore. The subject is contextualised for them.*

In addition, she seems to be aware of the exam task types in all the papers since she reported that it would be necessary to practise more multiple-choice tasks as they are required in the *Cambridge English: Key for Schools* exam and students are more used to True or False questions. She also pointed out that as candidates are examined for the Speaking paper in pairs, this kind of activity has been introduced in the preparation classes.

The teacher believes that what is taught in the English classes is aligned with what is assessed in the *Cambridge English: Key for Schools* exam and, for this reason, when preparing students for the exam the focus is more on the method and procedures and not on the content. Such alignment between course and exam has resulted in students feeling more confident and less worried about taking the exam. *In fact, to be very sincere it is quite natural . . . to sit for the exam . . . there is no big preparation . . . it is part of the daily strategies writing well, expressing oneself orally . . . in comprehension . . . so in general they [the students] do not worry to be exposed to an exam . . . it is very interesting . . . they go . . . ah . . . 'it is easy'.*

For the teacher, such positive attitudes to the exam is also because students begin taking Cambridge English exams when they are very young . . . *and I believe that because now there are several levels until FCE [First Certificate in English, now known as Cambridge English: First] . . . Before there was only FCE . . . then when they get there [FCE level] . . . they [students] are already prepared.*

Students who pass the exams feel very proud and the school acknowledges their achievement by hanging banners at the entrance of the building with their names, emphasising the fact that 100% of their students passed the Cambridge English exams.

Classroom assessment

Classroom assessment is based on a set of requirements such as attendance, quarterly and monthly tests which comprise writing tasks, grammar tests, listening tasks and extra class activities such as reader assignments, in which all the four skills are assessed.

All achievement tests are prepared by the teacher. According to her, there is not much concern about the *Cambridge English: Key for Schools* exams in the first assessment of the year in terms of task types and test procedures. For instance, transferring answers to the answer sheet is one aspect that students are trained for as they are not required to do this in regular tests.

The test paper analysed for this research was administered in the first quarter and consisted of two parts: reading comprehension and grammar. The reading part comprised four texts of different genres, three short stories and an email, each one with an average of 17 lines, followed by either six multiple-choice questions with three alternatives or six True or False questions. In this case, the false statement needed to be corrected. Most of the questions focused on checking text information. Only the email required the identification of the writer and of the interlocutor.

The grammar part consisted of gap-filling, rewriting sentences using target words, and multiple-choice gap-filling questions on the modals *must/mustn't*; pronouns *some, any, no*; *everybody, everywhere, everything*. In this part, there were also vocabulary activities in different formats (crossword puzzles, gap-filling and picture–word matching) about places around town and traffic signs. Both parts of the test used isolated sentences to check students' knowledge of vocabulary and grammar rules. The content of the lesson was in line with the content of the test and with the grammar content of the *Cambridge English: Key for Schools* exam.

Discussion

Table 1 brings together all the data gathered in an attempt to establish a parallel between *Cambridge English: Key for Schools* exam content/specifications and 1) the observed lessons; 2) the achievement test; 3) the coursebook and materials; and 4) the English language course programme. Not all the grammar content nor functions and notions covered by the exam, however, were listed in Table 1, just the aspects which are common to the exam and to the lessons, achievement test, course material and the programme.

Classroom observations showed that teaching follows the English course programme, which has many points of contact with the *Cambridge English: Key for Schools* exams in terms of: 1) skills: reading, listening, writing and speaking, allowing students to better understand and express ideas, and 2) view of language: English language is seen as a social and interactional activity.

In addition, the lessons involved a range of tasks and material resources for skills development like the coursebook and book assignments based on readers, DVD and YouTube listening activities, presentations, pairwork and student–teacher oral tasks, and narrative writing. One could conclude, therefore, that language is seen as a social activity approached through real communicative situations such as: understand and discuss how important house chores and house organisation are for a family; talk about a short story they have read; write a narrative text based on an episode of *Smallville*, an American TV series; understand a video showing friendship and love among adolescents.

The reading and writing genres assessed in *Cambridge English: Key for Schools* were either taught in class as seen in lesson observations or are likely to be taught at some point in the year since they are already part of the course programme.

The exam involves a wide range of tasks and test formats, and only some of them were identified in the lessons, achievement test, coursebook and the programme. The

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Table 1: Comparing the contents of Cambridge English: Key for Schools, lessons, achievement test, coursebook/materials and course programme

	Cambridge English: Key for Schools exam	Observed lessons	Achievement test	Coursebook/materials	Course programme
Reading genres	Notes, signs Dialogues, short texts	No Yes	No Yes	Yes Yes	Yes Yes
Writing genres	Form completion Guided short note, email, postcard	No No	No No	Yes Yes	No No
Task types	Matching, multiple choice, multiple-choice cloze, open cloze, gap-fill, true/false/does not say, and word completion	Matching, gap-fill, question answer	Multiple choice, true or false, gap-fill, sentence rewriting	Matching, gap-fill, sentence writing, question answer, multiple choice	Question answer, gap-fill
Speaking Patterns of interaction	Interlocutor–candidate Candidate–candidate	Teacher–student Student–student Student alone	There is no oral achievement test	Pairwork	Not specified
Grammar content	The grammatical content of the exam is the same as that found in the lessons, achievement tests, coursebook/materials and course programme For a full list of grammatical content covered, see the Handbook for teachers (Cambridge ESOL 2012:4–5)	Modals (must, mustn't), present continuous for present actions and future meaning, present simple, future with going to, quantitative pronouns, gerunds as subjects and objects	Modals (must, mustn't), quantitative pronouns	*For the 1st quarter of the school year:* Present simple, present continuous, past simple, past progressive, short and long comparative and superlative adjectives, must/mustn't, have to/don't have to, some/any/no/every, and be going to	*For the 1st quarter of the school year:* Past simple, past progressive, short and long comparative and superlative adjectives, must/mustn't, have to/don't have to, some/any/no/every, and be going to
Functions, notions and communicative tasks	The functions, notions and communicative tasks of the exam are the same as those found in lessons, achievement tests, coursebook/materials and course programme For a full list of functions, notions and communicative tasks covered, see the Handbook for teachers (Cambridge ESOL 2012:4)	Understanding and producing simple narratives, talking about future plans and intentions, giving warnings and stating prohibitions, expressing preferences, likes and dislikes	Not specified	*For the 1st quarter of the school year:* Talking about future plans or intentions, asking the way and giving directions, giving and responding to invitations, giving warnings and stating prohibitions, asking and giving/refusing permission to do something, expressing preferences	*For the 1st quarter of the school year:* Asking and giving directions, asking for permissions, making invitations

students are more used to matching, question–answer and gap-fill, which are part of the course programme. Multiple-choice questions are used extensively in the exam and, although not part of the programme, appear in the achievement test. The teacher acknowledges that introducing multiple-choice questions into the achievement test was an effect of the exam: *This . . . with multiple-choice is also a preparation . . . they didn't do it before . . . the usual is more true or false and justification . . . they are more used to this type . . .*

The development of speaking in class is done mostly by the teacher and the pattern of interaction is usually teacher–student or student–teacher in short interactions or in longer ones (individual presentations). The only student–student interaction observed was in a very controlled practice following a script presented in the coursebook which focused on expressing likes and dislikes. In this activity, students were set in pairs to express their likes and dislikes using sentences 'I hate doing . . .' or 'I like . . .' and their partner had to agree or disagree using the expressions 'So do I' or 'Really, I don't'. However, the exam requires candidates to interact first with the examiner and afterwards with each other using prompt material. The former pattern of interaction has more similarities to the one students have with the teacher when she asks open-ended questions to the group. In such interaction, however, students usually offer no more than a few sentences as the whole class participates and takes turns. This means that not all students can take part, but only the more outgoing and confident ones. Thus, in class, we did not see much of the pair interaction required in the exam when they have to ask and answer Wh-questions. Furthermore, it was observed that the teacher centralises the speaking activities instead of allowing students to interact more among themselves. The task types used in class for the development of speaking are not similar to the course programme, the coursebook or the *Cambridge English: Key for Schools* exam. In all of them, speaking is seen as an interactive activity involving pairwork. The course programme sets speaking objectives such as asking and giving directions, or inviting somebody to go out, while the coursebook presents pairwork activities such as asking your friend how often they do housework.

As far as grammatical content is concerned, it is possible to notice similarities between the exam and the lessons, course programme and coursebook. As both the coursebook and the *Cambridge English: Key for Schools* exam follow CEFR Level A2, such similarity was expected. In addition, as pointed out by the teacher, grammar content for the course programme and lessons was based on the coursebook. The same is true of the functions, notions and communicative tasks as there are many similarities between the observed lessons, coursebook, course programme and exam.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Having presented all the data, we now proceed to answering the research questions of this investigation:

What understanding of the Cambridge English: Key for Schools exams does the Brazilian teacher of English have? How does the teacher interpret the exam construct?

The data shows that the teacher interviewed is familiar with the levels of proficiency required by Cambridge English exams. For this reason she can make recommendations as to which students should take exams earlier or later. For instance, she can decide if a student who is in a grade when students usually take *Cambridge English: Preliminary for Schools* should take *Cambridge English: Key for Schools* instead or vice versa.

We could also conclude that she understands the exam construct and the view of language which underlies it. She understands that once the students are able to use language to express themselves and understand each other there is no need for intensive preparation for the *Cambridge English: Key for Schools* exam. She understands that the Speaking paper requires pair interactions and acknowledges that the lesson time in class is not enough for preparing students for this part of the exam. For this reason the school offers extra preparation classes for candidates in this skill.

We can conclude, therefore, that the teacher is familiar with the exam construct and with its key aspects such as levels of proficiency, skills, question types and situations assessed as well as with the underlying view of language.

Does the exam have any influence on the English taught as a subject at school? If so, what is the nature of this washback?

In our study, since the students who decide to take *Cambridge English: Key for Schools* have extra speaking lessons, the teacher does things she would not otherwise do, the students study harder, and, therefore, we could conclude that *Cambridge English: Key for Schools* exerts a general effect on student preparation and on teacher practice (Watanabe 2004). On the other hand, as described in Table 1, the multiple-choice task format was not frequently used in the lessons and students are not accustomed to it. In order to better prepare the students for the exam, this format was introduced in the achievement test as well as in the reading, writing and listening mock exam. Therefore, we could conclude the exam also exerts a specific effect on the teaching practice as she chooses task types which resemble the ones used in *Cambridge English: Key for Schools*.

Classroom observations were focused on regular lessons. In those lessons there was not a single reference to the exam or any practice directly related to the exam. In general the lessons followed the course programme, which does not make any references to Cambridge English exams either. When the observation started, we had no information regarding how *Cambridge English: Key for Schools* was presented to the students. The information that there was specific preparation for the exam was given by the teacher in the interview: *We already have a preparation for all students and for those who [are] enrolled [in] the exam we have a specific preparation especially because of the interview part*. Thus, it cannot be claimed that *Cambridge English: Key for Schools* determines everything that happens in class. At the same time, it affects the lessons, as our data shows. Therefore, we could conclude that although there is washback, it is not strong, which means that the exam affects some things that happen in the classroom but not everything.

Length of the effect is an aspect which could not be focused on in our study since it is not a longitudinal study. For Watanabe (2004), if the influence of the exam lasts after it is taken it can be considered long-term washback, but if it disappears as soon as it is over it is short term. In this school, they administer exams from *Cambridge English: Starters* (also known as *Young Learners English: Starters*) to *Cambridge English: First* level. Thus if students decide to take all of them during their elementary and secondary education it will mean taking an exam every one or two years. Consequently, the influence of one exam will probably continue until the next one. This situation reflects what the interviewed teacher referred to as getting accustomed to the exam: *When the students reach Cambridge English: Key for Schools they are already used to the exam*.

A fourth dimension of washback is intentionality, as exams could exert effects which are intended, but also unintended ones (Messick 1989). Intended effects are usually associated with positive washback and unintended ones with negative washback. This study shows that the teacher and school are very proud of the students' exam success. According to the teacher the last applicant who did not pass the exam was two years ago and this was considered an unusual case. However, the average number of applicants for *Cambridge English: Key for Schools* is 25% of the total 8th grade student population, which means that 75% of the students do not take the exam. It would be very interesting to find out why those students do not sit the exam. Do they feel intimidated or are afraid of being unsuccessful? Could this be a negative effect? Some questions can be raised from this situation. If the school is proud of their success rate are they discouraging students from sitting the exam who they are not completely sure will pass? Also, students who lack confidence but have a good level of English may feel intimidated because failing this exam would be an exception in this school. There may be other factors such as parents who do not want to pay the exam fee or who think their children should wait to reach a higher level. On the other hand, the exam is not mandatory and is not used for any screening purposes. Therefore, pressure is not so intense as if it is for a university entrance examination for example. Other than that, there was no evidence of negative washback, considering the scope of the investigation.

Considering the first three hypotheses presented by Alderson and Wall (1993) related to what and how teaching can be influenced, there is no evidence to conclude that the *Cambridge English: Key for Schools* exam influenced the teacher to change the content of classes or her approach to teaching from what was set by the course programme and coursebook. However, it has motivated the use of reading and listening mock exams taken from the teacher's digital coursebook which become part of the achievement tests that start in June and are taken by all of the students. The Speaking paper requirement of pair interaction, however, is dealt with in the preparatory extra lessons only for the students registered for the exam, not in regular classes. Therefore, there seems to be a moment during the year when the exam begins to trigger changes in teaching.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Regarding the next two hypotheses, that is, 4) a test will influence the rate and sequence of teaching, and 5) a test will influence the degree and depth of teaching, we can conclude, based on what the teacher said during the interview but not witnessed during our observations that there is a specific moment during the year when the exam samples will be administered. However, such administration is not predicted in the course programme. Thus it is possible to say that the exam influences the teaching practice but at the same time it does not affect the planning of the course programme.

All the aspects related to teacher attitudes and behaviour towards *Cambridge English: Key for Schools* seem to be positive as the interviewed teacher was part of the decision-making process to bring Cambridge English exams to the school. This brings us to the sixth hypothesis, that is, 6) a test will influence attitudes to the content, method, etc. of teaching. We could not identify school attempts to prepare, train or encourage teachers to offer better exam preparation or any pressure on them to reach high pass rates, or anything that might lead to negative attitudes towards teaching.

Alderson and Wall (1993) state that 7) tests that have important consequences will have washback and conversely, and 8) tests that do not have important consequences will have no washback. Cambridge English exams are not used for screening or for students' assessment in the school and, therefore, do not appear to be high stakes in this context. They are suggested to the students in order to foster the development of their proficiency in English and meet parents' expectations. Besides, the exam is not taken by all the students, and, therefore, while it can be of relevance to some because it brings confidence and motivation, it is not relevant to all of them, as the perception of relevance is always relative.

It is very common to see teachers relying heavily on materials to prepare students for the exam as discussed by Cheng (1997). However, in our study the influence of the Cambridge English exams on the coursebook and materials used in class was not identified. First of all, the coursebook chosen for the 8th grade is from Oxford University Press and in the Students' Pack there is no reference to *Cambridge English: Key for Schools*. In addition, according to the teacher interview, when selecting a coursebook the school did not include criteria related to the exam. However, the pack includes online access to the exam practice test and the version for the interactive white-board used by the teacher also offers exam practice. Thus we can conclude that although the exam does not exert influence directly on material selection, it plays an important role when the teacher is preparing students for the exam as the mock exams used in class are taken from the digital version of the book.

Conclusion

This research aimed to investigate if and how *Cambridge English: Key for Schools* influences the teaching practices in a private regular elementary school in São Paulo city. Thus its results are based on a case study of one teacher in a particular context and they may not be generalisable to other contexts.

We can conclude that although *Cambridge English: Key for Schools* and the teaching of English at the setting investigated had many aspects in common, there is no evidence to conclude that the exam exerts a strong effect on teaching, as the contents of the regular lessons or the teacher's approach seemed not to have been changed by the exam. Some influence, however, was observed.

A few aspects would be worth pursuing further. For instance, it would be interesting to investigate the effect of the exam on the students who register for the exam versus on the ones who do not register, as well as to understand the motivations of those who take the exam and of those who do not take it. As the participant teacher was also the English language co-ordinator, it would be interesting to investigate the other teachers' attitudes towards the exam and its effects on their practice. Another follow-up study would be to discuss with the teacher the results obtained and register her perceptions about them.

References

Alderson, C and Wall, D (1993) Does washback exist? *Applied Linguistics* 14 (2), 115–129.

Cambridge ESOL (2012) *Cambridge English: Key for Schools: Handbook for Teachers*, Cambridge: Cambridge ESOL.

Cheng, C (1997) How does washback influence teaching? Implications for Hong Kong, *Language and Education* 11 (1), 38–54.

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

Messick, S (1989) Validity, in Linn, R (Ed) *Educational Measurement*, New York: Macmillan, 13–103.

Watanabe, Y (2004) Methodology in washback studies, in Cheng, L and Watanabe, Y (Eds) with Curtis, C, *Washback in Language Testing*, New Jersey: Lawrence Erlbaum Associates, 19–36.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Young learner perceptions of test taking in Japan

JENNIFER M TOEWS-SHIMIZU SEIGAKUIN UNIVERSITY, AGEO, JAPAN

The purpose of this research was to investigate Japanese young learner perceptions of the *Cambridge English: Starters* examination as there is little research regarding young test taker attitudes and perspectives in Japan.

The study of positive effects which effective test design might have upon student perceptions of tests may be helpful in shedding light on improved student receptiveness to tests and possibly long-term positive effects on student motivation, study habits or in-class attentiveness. In this study, pre- and post-test attitudes towards *Starters* of grade 4 elementary students will be examined.

The *Starters* test is set at Common European Framework of Reference (CEFR) (Council of Europe 2001) Pre-A1 level, and is part of the *Cambridge English: Young Learners* (also known as *Cambridge Young Learners English (YLE)*) suite of exams which is designed to:

1. Sample relevant and meaningful language use.
2. Measure accurately and fairly.
3. Present a positive first impression of international tests.
4. Promote effective learning and teaching.
5. Encourage future learning and teaching.

(Cambridge ESOL 2012)

The *Starters* test layout is an easy-to-follow format and requires little explanation. Task images are colourful, clear and unambiguous, setting a 'safe' and non-trick-question-like place for students to perform. The *Cambridge English: Young Learners* Speaking standardisation material clearly sets the assessor's role as the provider of scaffolding when needed, in order to ensure that the test taker is supported and eventually (even with support) able to 'complete' the task at hand (Cambridge ESOL 2010). The idea is to leave the child with a positive test-taking experience, no matter what level they are, hence providing motivation to continue studying the language and attempt testing again.

Introduction

In general, both educators and students place high value on test results in Japan. Some say excessive concern with tests has become a problem on a national level (Rohlen 1983). Post-war systems of incentives and penalties have provided a basis for Japanese people to view standardised testing as a crucial aspect in personal achievement (Reedy 1999).

However, despite the importance of standardised testing in Japan, there is very little research on young learner test taker perspectives. One possible reason for this gap in the literature may be that until recently (about one year ago), foreign language studies for young learners were limited to private institutions. English as a foreign language was introduced as a mandatory subject in public schools in 2011 for grade 5 and 6 students (age 11 and 12). That said, informal conversations with private school elementary English teachers in Tokyo reveal that the discouragement of testing and assigning of grades to foreign language study seems to be held by some administrators and teachers (both Japanese and native English speakers) based on the idea that testing is a direct threat to the 'fun', 'likeability' and need for a 'mood of success' in the elementary English classroom. This is very different from the view held by junior high and high school language teachers. This may also serve as another reason for the lack of research into the testing perspectives of Japanese young learners, since it seems that not much formal testing is actually taking place in these classes.

This research explores the perceptions of young learners towards a reading and listening test. It will examine how test perceptions change from before to after a test is administered. Any changes in pre- and post-test perceptions will also be examined. Prevailing attitudes, commonalities and differences of test perceptions will also be revealed.

Research questions

1. How do young foreign language learners perceive tests?
2. How does test perception change from before the test is administered, to after the test is complete?
3. If student perceptions of testing change from negative to positive (or vice versa), how can this knowledge be used for future development of effective tests?
4. Are there any prevailing attitudes commonly experienced by students prior to test administration?
5. Are there any prevailing attitudes commonly experienced by students after a test is administered?

Hypothesis

1. Both typically 'high engagement/high performance' and 'low engagement/low performance' types of student attitudes will change in the pre- and post-test written responses.
2. The 'high engagement/high performance' student attitudes will show higher levels of positive and motivated characteristics than the 'low engagement/low performance' students in both the pre and post responses.
3. The 'low engagement/low performance' student attitudes will improve slightly from the pre- to the post-test responses.
4. It is expected that both types of participants, even those who expressed apprehension before the test, may express that the test was not as 'bad' as they had expected it would be.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Participants

The participants in this study were from the fourth year of a private, girls' elementary school in Tokyo. Six participants were chosen from two classes of 45 grade 4 students. Students had been randomly placed into the two classes of 45 at the beginning of the school year and test results from the current and previous term indicated equal levels of academic ability were represented in both classes.

Three of the participants represent typically 'high engagement/high performance' students (to be referred to as Group A from now on) while the other three participants represent the inverse (to be referred to as Group B from now on). Group A students are characterised by their high consistency of in-class, on-task behaviour, high motivation, positive attitudes and proficiency of study skills in the classroom and at home. The B participants represent the opposite type of student (consistent off-task behaviour, low motivation, negative attitudes and low proficiency in study skills). Homeroom teachers were asked to develop their own list of 6–10 students who represented the 'typically high engagement/high performance' and 'less than typically high engagement/high performance' students. Both groups showed consistent study, attitude and behaviour patterns across all their classroom subjects (i.e. maths, science, social studies, etc.).

Materials

The exam material came from the *Starters* 'Sample Paper A' which can be accessed on the Cambridge English website (www.teachers.cambridgeesol.org/ts/exams/younglearnersandforschools/ylestarters). The *Starters* exam has three papers covering listening, reading, writing and speaking. Each paper has between four or five parts which test different aspects of each skill. In this study, the focus is on Listening Part 1 and Reading Part 2. Classroom materials were prepared by the teacher as described in the following section.

Method

The exam administration and the events leading up to the exam should be specially noted due to differences in pre-exam practice tasks, class materials and exam administration in relation to the official administration methods of *Starters*.

Listening Part 1 requires students to listen and indicate where an object ought to be placed in a picture. This task tests students' knowledge of the nouns and prepositions found on the *Starters* vocabulary list (Cambridge ESOL 2012). In preparation for this exam, the teacher made up tasks very similar to the task found in the exam (i.e. 'Sample Paper A') which contained an image of a living room. One example of such a task (four lessons into the unit) involved a large poster of a house layout containing furniture being placed on the blackboard. On smaller magnet cards, pictures of various household objects were placed around the perimeter of the poster. During the warm-up, the class listened, as a group came to a consensus to choose the correct card and then an individual student (representing the group decision), placed it in the correct location in the house. Various tasks including pairwork ensued. Songs using prepositions of location were also included in each lesson. The last task in the fourth lesson included individual students making up 'silly sentences' using prepositions of location, after which the teachers listened and then placed the household items in the locations on the poster as stated by the student. It should be noted that the model language used in these classroom tasks was planned carefully after examining the *Starters* vocabulary list and the exam task.

In Reading Part 2, students read sentences that both describe a picture correctly and incorrectly and they must indicate whether the sentence is true or false by writing 'yes' or 'no' next to it. The same principle was applied to preparing students for this part of the exam as was used for preparing them for the Listening paper.

Students were prepared for the test over a 3-week period. Lessons consisted of songs, chorus speaking, flashcards, listening tasks, meaning-focused tasks and closed writing tasks focused on using prepositions of location. The students never completed any *Starters* practice tests.

The administration of the exam differed significantly from the official method. The students were only tested on Listening Part 1 and Reading Part 2. The listening task was completed by listening to the CD and the students writing their answers in a blue pencil. After this (because the assistant teacher feared the students would not understand the British English on the CD), the teachers read the script, resulting in a third opportunity to listen, and the students made any additions or corrections using a red pencil. The reasons for using two coloured pencils was so that the data from the CD listening could be compared to the teacher reading.

Data analysis

Three pieces of data were collected on each participant; classroom field notes, pre-test response comments and post-test response comments. The data was analysed for patterns, similarities, differences and overall tendencies.

The outcomes of this research aim to provide some insight into the perceptions of two opposite types of student. Any changes in perception may shed light on previously formed beliefs about testing as well as the impact of the *Starters* test on the test taker's attitudes.

Procedure

This study was carried out in four steps. Firstly, the classroom observations of the participants were made over a 3–4-week period. Field notes were made on their observed behaviours during classes (e.g. on-task behaviour, completion of homework, attitude to tasks, performance, etc.) in order to confirm patterns of student behaviours in each group. During these weeks, a unit on prepositions of location was taught. A test was announced about a week prior to the test date. It should be noted that though the students had not completed *Starters* practice tests, they completed teacher-prepared materials and tasks throughout the unit of study that were based on the format and content of the *Starters* test.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Secondly, on the test date, following 8 minutes of review of the test materials through flashcards, chorus speaking and a song, each student was given a blank piece of paper upon which they were asked to write their name, class and student number. The students were told (in Japanese) that the teachers were interested in knowing what they were thinking about prior to the start of the test, what kind of questions they might have in their minds, what kind of feelings they might be experiencing and any connected or disconnected thoughts they might be experiencing just prior to the test they were about to take. The students were told that they could write anything on the paper, and they shouldn't feel influenced by how much others around them were writing. They were being requested to write only what was on their own mind and since they would be receiving no grade for this bit of writing, they should feel very free about the content and length of their writing. Students were also told that there is no one correct type of comment so that they would record their thoughts without fear of negative evaluation. No time limit was in place during the writing of pre- and post-test comments. The students were given about 6 minutes to write their comments (that is when the last student stopped writing). When all students had finished, the students were instructed to put the paper face down on the side of their desk to maintain their privacy during the test since they would need the paper again later.

Thirdly, a listening and reading test (using *Starters* Listening Part 1 and Reading Part 2) was conducted. It should be noted that this was not a live administration of the exam and procedures differed from those that would occur during an official administration of the exam. In particular they were given the opportunity to listen three times as opposed to the standard two times. The test was handed out in silence and then a brief explanation was given regarding the use of the British English which they would hear on the CD. The students were told that during the listening part of the test, they would first listen twice to the CD containing British English and then the exact content of the CD would be heard again, as read by the two teachers (using the script). They were instructed that during the first and second listening (i.e. the Cambridge English CD), they were to write down their answers using their regular pencil. During the third listening (as read by the two teachers), they were to mark or make changes to their answers using a blue pencil (therefore data reflecting their comprehension of the CD could be differentiated from their comprehension of their teachers' reading of the script).

Finally, after the test was finished, students were asked to turn their test papers face down and write post-test comments regarding their impressions, feelings, and questions, etc. on the bottom half of the paper upon which their pre-test comments had been recorded. The same instructions were given regarding how they should proceed with writing their comments. These instructions were again given in the L1. The last student finished writing her comments after about 7 minutes, upon which the test papers and the comments were collected.

After all the tests and comments were collected, the tension in the room was fairly high. Both teachers sensed a need to allow the students to debrief verbally in order to facilitate a time for students to share their response to this testing session with their friends and teachers. Pent-up feelings and uncommunicated misconceptions can lead to discomfort/dissatisfaction/mistrust on behalf of students and parents. During the audio-recorded 10-minute session, students openly and enthusiastically shared how they felt throughout the testing session. The teachers also took a casual survey (students raising hands to be counted) during this session, asking students how they felt about various aspects of the test. The sessions ended with many students smiling and looking relaxed. The tension seemed to have dissipated during the verbal debriefing session.

Results and discussion

Field notes – participant observations

Due to an over-estimation of teacher/researcher ability to both manage classroom teaching/tasks (plus other factors such as nosebleeds, etc.) *and* count/record 'exact' quantitative data on each student's behaviour, the teachers met after the classes to discuss general observations made on each participant over the 3-week observation period prior to administrating the test. The following notes on each student represent the observations made by both teachers over this period.

> Participant A 28, A 26 and A 11 were recommended by homeroom teachers based on high/perfect test performance in all courses of study, good communication skills, co-operation with others, high enthusiasm and motivation. Participant B 6, B 15 and B 20 were all recommended by homeroom teachers based on low test performance in all courses of study, poor attitudes, poor co-operation, low participation and motivation.
>
> Participant A 28 and A 26 showed perfect test performance (over nine months of classes), enthusiastic participation in lessons (raising hand, singing, speaking in loud voice, cheerfulness, always bringing English texts and supplies to class). These participants were not observed being off-task, chatting or complaining by either teacher.
>
> Participant A 11 showed perfect test performance (over nine months of classes) enthusiastic participation in lessons (raising hand, singing, speaking in a loud voice, cheerful). This participant did, however, fail to bring her textbooks and supplies to class on seven occasions during a 9-month period. Participant A 11 showed particular assertiveness in completing tasks beyond the expectations of the teacher, often expanding role-plays to include ad lib and 'play-language'.
>
> Participant B 6 showed below average test performance (over nine months of classes), low enthusiasm (never raises hand, doesn't sing, speaks in a low and mumbled voice, voices negative comments about lesson tasks, and forgot to bring her textbooks/supplies to class 17 times over a period of nine months). This participant was observed being frequently off-task, chatting and complaining by both teachers.
>
> Participant B 15 showed exceptionally low test performance (over nine months of classes), low enthusiasm (never raises hand, doesn't sing, speaks in a low voice, voices negative comments about lesson tasks, and forgot to bring her textbooks/supplies to class 19 times over a period of nine months). This participant was observed being frequently off-task, chatting and complaining by both teachers.
>
> Participant B 20 showed below average test performance (over nine months of classes), low enthusiasm (rarely raises hand, sometimes sings, speaks in a low and mumbled voice, voices negative comments about lesson tasks). This participant was observed being frequently off-task, chatting and sometimes complaining by both teachers.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Pre- and post-test comment results

Table 1 represents the translated (originally written in Japanese) pre- and post-test comments which the students wrote. It is observed that the Group A word count (Mean (M) = 24.5) exceeded the Group B word count (M = 15.25). Participant A 26 also included a sketch below her post-test comments of a little girl whose heart is pounding.

As can be observed at a glance, the comments for Group B are much shorter and fewer in number than Group A. It could be concluded that Group B participants who frequently demonstrated a lack of interest in their studies also showed less ability or interest in reflecting upon their work or performance. Therefore the difference in comment length could be linked to their lack of engagement in class.

Pre- and post-test comment analysis

In order to analyse the pre- and post-test comments, categories of comment types were developed and the comments categorised accordingly. Each category, its explanation and examples of the comments will be listed below. The category types are as follows; negative, positive, expressions of confidence, lack of confidence, expressions of enjoyment and anticipation, and negative anxiety.

Negative comments represent comments which reflected low motivation, lack of enjoyment, negative anxiety, lack of confidence in ability, lack of understanding of test/studied content and expression of difficulty. Some examples of these types of comments are as follows: 'I am very nervous', 'There are some parts that I am not good at', and 'I'm bored'. As can be seen in Table 2 (pre-test comments), Group A (M = 1.00, Standard Deviation (SD) = 1.73) and B (M = 1.00, SD = 0.00) make the same number of negative comments, though this is not significant (t = 0.00, p = 1.00). However, in Group A, only one student is making these comments while in Group B, all three students report negative feelings. Again, in the post-test comments (Table 3), both groups make two negative comments each, but in Group A (M = 1.00, SD = 0.94) only one student makes negative comments, while in Group B (M = 1.00, SD = 0.47), two of the students report negatively. This is not significant (t = 0.00, p = 1.00).

Positive comments represent comments which reflected high motivation, enjoyment/anticipation, lack of anxiety, confidence, understanding of test/studied content and expressions of the easiness of the test/studied materials. Some examples of these statements include: 'I'm happy', 'It seems fun', 'I will do my best' and 'I love English tests'. In the pre-test comments, Group A made 11 positive comments (M = 3.67, SD = 0.47) while Group B (M = 2.00, SD = 1.25) students made only four. This figure is not significant (t = 2.17, p = 0.10). In the post-test comments, Group A (M = 1.67, SD = 0.47) made five positive comments while Group B

Table 1: Pre- and post-test comments

		Pre-test comments	Post-test comments
Group A	A 28	I'm happy, looking forward to the test, a little bit sleepy. I think I can do it. I am feeling calm.	It was interesting. There was a little happening with my pencil but it was okay in the end.
	A 26	I am really nervous. There are some parts that I am not good at and I don't understand at all. But I studied at home so I will believe in myself and do my best!	The CD was difficult to understand. When the teacher read it at the third time, it was easy to understand and I felt very satisfied. (The girl in her picture is saying 'Doki-Doki' which means, 'my heart is pounding'.)
	A 11	I'm not so worried. It seems fun. But maybe I should be a little nervous (have tension) when I'm taking the test. I love English tests! I wonder what kind of questions there are.	I am now really hungry. Also, the test was really good! I feel good!
Group B	B 6	I'm bored.	So-so easy. (It was kind of easy.)
	B 15	I am very nervous. I want to get a full score.	There were some questions I didn't understand, but I did my very best.
	B 20	I'm a little nervous but it seems fun and somehow I'm happy. I will do my best.	I am hungry. It was fun. I want to do it again because it was fun.

Table 2: Pre-test comment analysis

Comment type/topic	Group	Pre-test number of comments	No. of students making comments	Mean	SD	t	p
Negative	Group A	3	1	1.00	1.73	0.00	1.00
	Group B	3	3	1.00	0.00		
Positive	Group A	11	3	3.67	0.47	2.17	0.10
	Group B	4	2	2.00	1.25		
Expression of confidence	Group A	2	2	0.67	0.47	2.47	0.07
	Group B	0	0	0.00	0.00		
Lack of confidence	Group A	2	1	0.67	0.94	1.23	0.28
	Group B	0	0	0.00	0.00		
Enjoyment & anticipation	Group A	5	3	1.67	0.47	0.78	0.48
	Group B	3	1	1.00	1.41		
Negative anxiety	Group A	1	1	0.33	0.67	1.42	0.23
	Group B	2	2	1.00	0.47		

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Table 3: Post-test comment analysis

Comment type/topic	Group	Pre-test number of comments	No. of students making comments	Mean	SD	t	p
Negative	Group A	2	1	1.00	0.94	0.00	1.00
	Group B	2	2	1.00	0.47		
Positive	Group A	5	3	1.67	0.47	0.43	0.69
	Group B	4	2	2.00	1.25		
Expressions of confidence	Group A	3	2	1.50	0.82	3.17	0.03
	Group B	0	0	0.00	0.00		
Lack of confidence	Group A	1	1	0.33	0.47	0.00	1.00
	Group B	1	1	0.33	0.47		
Enjoyment & anticipation	Group A	4	3	1.33	0.47	0.00	1.00
	Group B	4	2	1.33	1.24		
Negative anxiety	Group A	1	1	0.33	0.47	1.22	0.29
	Group B	0	0	0.00	0.00		

(M = 2.00, SD = 1.25) made four. This is not significant (t = 0.43, p = 0.69).

Expressions of confidence are represented by comments such as: 'I think I can do it', or 'I studied at home so I will believe in myself'. During the pre-test comments, Group A (M = 0.67, SD = 0.47) made two expressions of confidence while Group B (M = .00, SD = .00) made no expressions of confidence. This is not significant (t = 2.47, p = 0.07). In the post-test comments, Group A (M = 1.50, SD = 0.82) made three expressions of confidence regarding the test while Group B (M = 0.00, SD = 0.00) made no expressions of confidence. This figure is somewhat significant (t = 3.17, p = 0.03).

Comments expressing lack of confidence referred to lack of certainty in the self and lack of confidence in the understanding of the test/studied content. Some examples of these types of comments include: 'there are some parts I'm not good at' and '[some parts] I don't understand at all'. In the pre-test comments, Group A (M = 0.67, SD = 0.94) made two comments expressing lack of confidence while Group B (M = 0.00, SD = 0.00) made no comments. This figure is not significant (t = 1.23, p = 0.28). In the post-test comments, Group A (M = 0.33, SD = 0.47) made one comment expressing lack of confidence while Group B (M = 0.33, SD = 0.47) also made one comment. This figure is not significant (t = 0.00, p = 1.00).

Expressions of enjoyment and anticipation included references to happiness, satisfaction and feeling good. Some examples include: 'It seems fun, somehow I'm happy, I will do my best, and I wonder what kind of questions there are'. In the pre-test comments, Group A (M = 1.67, SD = 0.47) made five comments expressing enjoyment or anticipation of the test while Group B (M = 1.00, SD = 1.41) made three comments. This figure is not significant (t = 0.78, p = 0.48). In the post-test comments, Group A (M = 1.33, SD = 0.47) made four comments while Group B (M = 1.33, SD = 1.24) also made four comments. This figure is not significant (t = 0.00, p = 1.00).

Negative anxiety comments referred to nervousness, uneasiness or tenseness expressed by the student. Some examples include: 'I am really nervous', picture of a child with their heart pounding and 'I am a little nervous'. In the pre-test comments, Group A (M = 0.33, SD = 0.67) made one reference to anxiety while Group B (M = 1.00, SD = 0.47) made two comments regarding anxiety. This figure is not significant (t = 1.42, p = 0.23). In the post-test comments, Group A (M = 0.33, SD = 0.47) again made one comment on their anxiety while Group B (M = 0.00, SD = 0.00) made no comments. This figure is not significant (t = 1.22, p = 0.29).

Table 4 presents a simple comparison of pre- and post-test comment types of Group A and B. It is observed that both groups made the same number of negative comments in both pre and post comments. Both groups made more negative comments before the test than after. While Group A made 11 pre-test positive comments, Group B only made four. In the post comments, Group A made five and Group B made four. In fact, Group B made the same number of positive comments

Table 4: Pre- and post-test comment type comparison

Comment type/topic	Group	Pre-test number of comments	Mean	Post-test number of comments	Mean
Negative	Group A	3	3	2	2
	Group B	3		2	
Positive	Group A	11	7.5	5	4.5
	Group B	4		4	
Expression of confidence	Group A	2	1	3	1.5
	Group B	0		0	
Lack of confidence	Group A	2	1	1	1
	Group B	0		1	
Enjoyment & anticipation	Group A	5	4	4	4
	Group B	3		4	
Negative anxiety	Group A	1	1.5	1	0.5
	Group B	2		0	

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

before and after the test. Group A made two comments expressing confidence and three in post comments while Group B made no comments expressing confidence. With regard to comments indicating lack of confidence, Group A made two comments in pre and one in post while Group B made no pre comments but one in the post comments. Group A expressed five comments of enjoyment and anticipation in the pre-test comments while Group B made only three. Group A made one less comment (four) while Group B increased their comments of enjoyment to four. This could indicate that even though Group B showed lower performance scores, the test itself was more positive than expected (hence the increase in enjoyment levels). With regard to the expression of negative anxiety, Group B made two pre-test comments on negative anxiety and none post.

When looking at the overall comparison of the number of comments produced by both groups, some tendencies and differences may be observed. The two groups' combined Means for each type of comment will be compared. With regard to negative comments, both Group A and B report a decrease in negative comments (pre M = 3, post M = 2) after the test. This could indicate that both groups realised that the test was not as 'bad' or unpleasant as they thought it was going to be. This could be interpreted as a positive testing experience for both types of students regardless of actual performance, motivation or lack thereof, study habits, etc. It is also interesting to note that when comparing the number of positive and negative comments reported by both groups, both groups report more than double the number of positive comments (pre M = 7.5, post M = 4.5) as negative comments (pre M = 3, post M = 2). This again indicates that both types of students had more positive perceptions than negative perceptions of the testing session. This clearly indicates that, again, regardless of the student's individual motivations, interests, performance level, etc., the test was still a positive and pleasant experience. This reflects positively on the test. With regard to the expression of confidence, Group A alone reports confidence before the test and shows an increase (pre M = 0.67, post M = 1.50) in confidence after the test. This could indicate that students who are engaged and motivated have high performance levels, etc., gained more confidence through the act of taking this test. On the other hand, students who are less engaged or motivated like those in Group B indicate no change or increase in their confidence (pre/post M = 0.00). It is interesting to note that both groups report the same number (M = 4) of pre and post comments on enjoyment and anticipation. Lastly, overall, the number of negative anxiety comments decreased (pre M = 1.5, post M = 0.5) between the pre- to post-test comments. It is interesting to note, however, that the same student (A 26) is reporting the negative anxiety in both the pre and post situations (see Table 1). This student may show consistent foreign language anxiety no matter what the situation is. Group A reports the same number of negative anxiety comments between pre/post test (M = 0.33) while Group B, on the other hand reports a decrease of anxiety in their post-test comments (pre M = 1.00, post M = 0.00). In this case, two students reported anxiety in the pre-test situation (B 15, B 20), with neither reporting anxiety in their post-test comments (see Table 1). This shows a reduction of anxiety after the test was completed.

Combined Reading and CD Listening test performance results

Table 5 holds the combined scores of the Reading and Listening test. Table 5 represents the Listening scores of the answers derived from listening to the Cambridge English CD only. Group A performed better than Group B (t = 4.62, p = 0.01). This figure is significant.

Table 5: Combined Reading and Listening test performance (CD Listening)

	Participant	CD Listening and Reading score*	Group mean	SD	t	p
Group A	A 28	10	10	0.00	4.62	0.01
	A 26	10				
	A 11	10				
Group B	B 6	7	6	1.50		
	B 15	4				
	B 20	7				
Overall mean		8				

*The highest possible test score was 10.

Reading and Listening (CD and read by teacher) score results

Table 6 shows the combined scores of the Reading and Listening test. However, Table 6 represents the Listening scores of the answers derived from listening to the Cambridge English CD *as well as* the reading of the script by the two teachers. Group A (M = 10, SD = 0.00) scored a higher mean than Group B (M = 6, SD = 1.62). This figure is somewhat significant (t = 4.50, p = 0.01). When comparing the scores of Table 5 and Table 6, no difference in scores is observed, which might suggest that the assistant teacher's discomfort with spontaneously using a non-familiar accent is perhaps unfounded.

Table 6: Combined Reading and Listening test (Listening: CD and teacher reading)

	Participant	CD/teacher Listening and Reading score*	Group mean	SD	t	p
Group A	A 28	10	10	0.00	4.62	0.01
	A 26	10				
	A 11	10				
Group B	B 6	7	6	1.50		
	B 15	4				
	B 20	7				
Overall mean		8				

*The highest possible score was 10.

Listening performance

As seen in Table 7, Group A's performance (M = 5, SD = 0.00) is superior to that of Group B (M = 3, SD = 1.00). This figure is somewhat significant (t = 3.46, p = 0.03).

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Table 7: CD Listening scores

	Participant	CD *	Group mean	SD	t	p
Group A	A 28	5				
	A 26	5	5	0.00		
	A 11	5				
Group B	B 6	3			3.46	0.03
	B 15	2	3	1.00		
	B 20	4				
Overall mean		4				

*The highest possible score was 5.

Table 8 shows the Listening scores derived from listening to both the Cambridge English CD and the reading of the script by the teachers. Group A performed better than Group B. The figure is somewhat significant (t = 3.46, p = 0.03).

Table 8: CD and teacher Listening scores

	Participant	CD and teacher*	Group mean	SD	t	p
Group A	A 28	5				
	A 26	5	5	0.00		
	A 11	5				
Group B	B 6	3			3.46	0.03
	B 15	2	3	1.00		
	B 20	4				
Overall mean		4				

*The highest possible test score was 5.

The Reading scores, as seen in Table 9, reflect a higher mean in Group A than B. This figure is somewhat significant (t = 3.46, p = 0.03).

Table 9: Reading scores

	Participant	Reading score *	Overall mean	SD	t	p
Group A	A 28	5				
	A 26	5	5	0.00		
	A 11	5				
Group B	B 6	4			3.46	0.03
	B 15	2	3	1.00		
	B 20	3				
Overall mean		4				

*The highest possible test score was 5.

Conclusion

In reference to limitations of this study, it ought to be noted that these students completed the *Starters* exam on fairly unfamiliar terms. While the content of the exam was integrated into classroom materials prepared by the teacher in preparation for the test, they never did any 'practice' tests. Within that framework, the classroom review of materials for listening also involved 'listening' to the teacher as opposed to listening to the *Starters* test CD, which the students encountered for the first time on test day. The lack of familiarity with the accents on the CD, however, appear not to have had as much of a negative effect as was expected. It should also be pointed out that this study involved a small number of participants in a particular context, making the results difficult to generalise to a larger population or other contexts.

For easy referral, the research questions and hypothesis will be restated here. It is to be noted that few findings or tendencies in this study are statistically significant.

Research questions

1. How do young foreign language learners perceive tests?
2. How does test perception change from before the test is administered, to after the test is complete?
3. If student perceptions of testing change from negative to positive (or vice versa), how can this knowledge be used for future development of effective tests?
4. Are there any prevailing attitudes commonly experienced by students prior to test administration?
5. Are there any prevailing attitudes commonly experienced by students after a test is administered?

With reference to question number 1, it seems that both types of students experienced some level of negative feelings, lack of confidence, negative anxiety as well as positive feelings, and enjoyment/anticipation before/after the test (though not statistically significant). What is interesting (though not statistically significant), is that Group A students' overall pre-test experience seems to have been much more positive (though not statistically significant) than that of the Group B students. We might argue that repeated positive test experiences could lead to higher motivation, better performance and possibly higher engagement during lessons.

The lack of comments expressing confidence by Group B in both the pre and post (p = 0.03) test shows a tendency of Group B students to experience no or little confidence before or after a test. Their lack of confidence is confirmed in their lower performance levels (t = 4.62, p = 0.01) in both reading and listening as seen in Table 5. Other conclusions regarding how students perceived the test in general are difficult to draw as only the lack of confidence figures were significant.

Research question 2 refers to any changes in perceptions from the pre to the post stage of the test. Though not significant, it seems that both types of students experienced a reduction of positive perceptions of the test when comparing the pre and post mean number of comments. During the typical post-test debriefing where the teacher/researcher usually takes a casual verbal survey, it is interesting to note that out of the 90 students, 42 students indicated (by hand-raising) that this test was more difficult than the usual tests they experienced in English class. So, it is possible that because they may have found the test more difficult, their positive perceptions which they experienced prior to the test were dampened.

If the mean figures of Table 4 are considered, it seems that though there are some slight changes in pre- and post-test perceptions they are only slight. So, it seems difficult to say

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

that there were any changes in the overall perceptions of either group of students. With that in mind, research question number 3 seems to then become insignificant.

Research questions 4 and 5 refer to any prevailing tendencies in either pre- or post-test perceptions. Again, since only one significant figure can be observed, we can conclude that less engaged and motivated students are not as likely to express any confidence in their performance or their preparation. In this experiment, Group B students made no comments expressing confidence either before or after the test was completed. However, only the post-test figure was significant when compared to Group A. We can duly understand that, as the performance levels of Group B were, in fact, lower than Group A, they would be less likely to feel confident about the test based on previous experiences.

Hypothesis

1. Both Group A and B attitudes will change in the pre- and post-test written response.
2. The 'high engagement/high performance' student attitudes will show higher levels of positive and motivated characteristics than the 'low engagement/low performance' students in both the pre and post responses.
3. The 'low engagement/low performance' student attitudes will improve slightly from the pre- to the post-test responses.
4. It is expected that both types of participants, even those who expressed apprehension before the test, may express that the test was not as 'bad' as they had expected it would be.

No strong support can be found in support of hypothesis statement 1 or 2. It seems that hypothesis statement 3 is fully supported in that Group B students' expressions of anxiety decreased (to no post-test anxiety), statements of enjoyment/anticipation increased and negative anxiety statements decreased after the test was completed. However, no figures were significant. It might be useful to explore this possible tendency in future studies.

Hypothesis 4 does not seem to be supported outright. However, as stated above, it seems that some of Group B participants' post-test comments turn towards a positive direction.

This study arguably contributes information for a particular context by confirming that typically 'high engagement/high performance' students are more likely to perceive tests in a more positive way than other types of students. Also, the 'low engagement/low performance' type of student is less likely to experience confidence before and after they have completed a test. Further research is needed to investigate young learner perceptions of tests in order to come to a better understanding of their attitudes.

References and further reading

Brown, J D and Yamashita S (Eds) (1995) *Language Testing in Japan*, Tokyo: The Japan Association for Language Teaching.

Cambridge ESOL (2010) *Instructions to Speaking Examiners: YLE Young Learners English Tests*, Cambridge: Cambridge ESOL, internal document.

Cambridge ESOL (2012) *Cambridge English Young Learners: Handbook for Teachers*, Cambridge: Cambridge ESOL.

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

MacIntyre, D and Gardner, R C (1991) Methods and results in the study of foreign language anxiety: a review of the literature, *Language Learning* 41 (1), 283–305.

Nottelmann, E and Hill, K (1977) Test anxiety and off-task behavior in evaluative situations, *Child Development* 48, 225–231.

Onwuegbuzie, A, Bailey, P and Daley, C E (2002) The role of foreign language anxiety and students' expectations in foreign language learning, *Research in the Schools* 9, 33–50.

Parkinson, B and Howell-Richardson, C (1990) Learner diaries, in Brumfit, C and Mitchel, R (Eds), *Research in the Language Classroom. ELT Documents 133*, London: English Publications and The British Council, 128–140.

Prins, J M, Groot, M and Hanewald, G (1994) Cognition in test-anxious children: The role of on-task and copying cognition reconsidered, *Journal of Consulting and Clinical Psychology* 62, 404–409.

Reedy, S (1999) *Recognizing and responding to cultural learning styles*, paper presented at the International TESOL Conference, New York City, March 1999.

Robinson, P (2001) Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition, *Second Language Research* 17, 368–392.

Rohlen, T (1983) *Japan's High Schools*, Berkeley: University of California Press.

Wigfield, A and Eccles, J S (1989) Test anxiety in elementary and secondary school students, *Educational Psychologist* 24, 159–186.

Wine, J (1971) Test anxiety and direction of attention, *Psychological Bulletin* 76, 92–104.

Woodrow, L (2006) Anxiety and speaking English as a second language, *RELC Journal* 37, 308–328.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Stakeholders' perceptions of the Occupational English Test (OET): An exploratory study

IVANA VIDAKOVIĆ RESEARCH AND VALIDATION GROUP, CAMBRIDGE ENGLISH LANGUAGE ASSESSMENT
HANAN KHALIFA RESEARCH AND VALIDATION GROUP, CAMBRIDGE ENGLISH LANGUAGE ASSESSMENT

Context of the study

The purpose of this exploratory study is to seek stakeholders' perceptions with regard to the appropriateness of the Occupational English Test (OET) to the healthcare sector, and to examine whether OET is an indicator of workplace readiness in terms of language proficiency. Stakeholders in this study consist of: OET test takers, teachers who prepare candidates for OET, employers and healthcare regulatory bodies who decide on the provisional registration of healthcare professionals. The study is exploratory in nature as it will be complemented with a cross-design where performance on OET can be compared to performance on other language exams used for the same purpose.

Description of OET

OET falls under the category of English for Specific Purposes (ESP). The test is designed to meet the specific needs of the healthcare sector in 12 healthcare professions, namely, dentistry, dietetics, medicine, nursing, occupational therapy, optometry, pharmacy, physiotherapy, podiatry, radiography, speech pathology and veterinary science. It assesses the English language competence of healthcare professionals who have gained their qualifications and training outside an English-speaking country.

It assesses four language skills: reading, listening, speaking and writing. The Speaking and Writing sub-tests are tailored to each profession, unlike the Listening and Reading sub-tests. While the latter are firmly grounded in the healthcare domain, in terms of topics, content, language and some tasks (e.g. note-taking while listening), they are not as profession specific as the productive skills tests. The largest groups of test takers are nurses (47%), doctors (23%), dentists (20%), and pharmacists (6%)[1].

Recognised by 30 regulatory healthcare bodies in Australia, New Zealand and Singapore, the Australian Department for Immigration and Citizenship (DIAC), universities and education providers (see Appendix 1), OET is currently available up to 10 times a year in 20 countries around the world (see Appendix 2 for a full list of countries and cities). The largest candidature is in Australia. Originally designed by Professor Tim McNamara of the University of Melbourne under contract to the Australian Federal Government in the late 1980s, OET is now owned by Cambridge Boxhill Language Assessment, a joint venture between Cambridge English and the Box Hill Institute in Melbourne, Australia. This recently formed strategic alliance is 'aimed at increasing the availability of the test to candidates in the UK, Europe and North America' (Cambridge English Language Assessment 2013), specifically in the countries with demand for foreign healthcare workers.

Research questions and design

The key research questions are:

a) What is the intended impact of using an ESP test, namely OET, when assessing the language ability of healthcare professionals?
b) To what extent is OET an appropriate language examination for the health sector in terms of its construct validity (i.e. content, skills/abilities assessed, format)?
c) To what extent are OET test takers perceived as ready for the workplace in terms of their English language ability and their confidence in using English in a healthcare context?

Qualitative and quantitative data were simultaneously collected in a mixed methods research design (MMRD). The analysis of each data strand was carried out independent of the other but when interpreting the results information was drawn from both strands. This approach enabled building a rich picture and the triangulation of information derived from multiple data sources enhanced confidence in the findings (see Greene, Caracelli and Graham (1989) for a discussion on reasons for mixing methods). Qualitative data collection instruments comprised interviews with representatives of healthcare regulatory bodies in Australia. Surveys were used in the quantitative analyses. These were administered to OET test takers, to teachers who prepare candidates prior to sitting for OET and to healthcare professionals.

Accordingly, the study sample included: (a) 585 OET test takers who currently work in private practices, hospitals, health centres, or research institutes – the majority of test takers work in Australia and New Zealand; (b) 27 OET teachers – the majority of whom work in Australia and New Zealand; (c) 40 healthcare professionals working in Australia – mostly medical doctors and a few pharmacists; and (d) two representatives of the healthcare regulatory bodies in Australia.

Key findings

Intended impact of using an ESP examination

Being intended for 'people who have trained as medical and health professionals in one country and wish to gain

[1] Based on January 2008–May 2013 data.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

provisional registration to practise their profession in an English speaking context' (The University of Melbourne 2009:1), it is to be expected that the OET would be taken by those who wish to gain professional registration and find a job in the healthcare sector. A vast majority of the respondents in the present study have indeed taken the test for those two tightly linked purposes (see Table 1). To a small extent, OET also appears to be used for university admission, as reported by 1% of the respondents, as well as for immigration or application for permanent residency in Australia.

Table 1: Reasons for taking the OET

I have taken the OET ... (more than one option can be selected)	Number of responses
... for the purposes of professional registration	540 (78%)
... for employment purposes	121 (18%)
... to enrol at a university	10 (1%)
Other (immigration or permanent residency application)	17 (2%)
Total	**688 (99%)***

*Percentages do not sum to 100 due to rounding.

Thirty-two percent of the test takers who responded believe that, as a result of preparing for OET, their ability to use English in their health-related context has improved a lot; 29% believe that they have experienced a moderate improvement; 16% believe that they have experienced a slight improvement while 19% believe they have not improved at all. However, there was a marked increase in their perception of confidence gained as a result of preparing for OET, as Figure 1 shows.

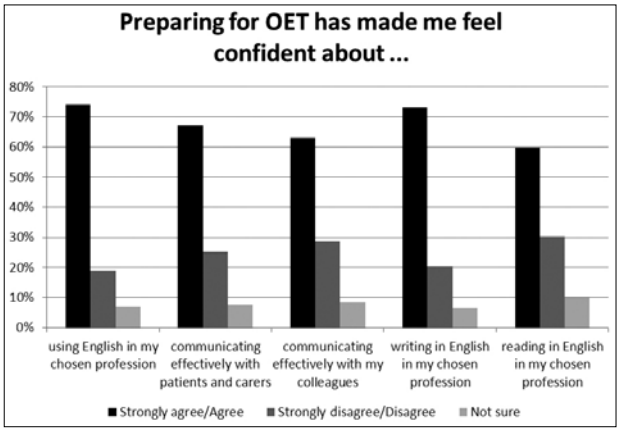


Figure 1: Confidence after preparing for the OET

Appropriateness to healthcare context

Both OET test takers and teachers believe that the major strength of OET is its relevance for the healthcare sector in terms of topics, language, tasks, scenarios and the language ability/skills its tasks require. The relevance of OET for the healthcare sector has a positive impact on the test takers because: a) preparation for OET prepares the test takers for performing language-mediated tasks in their chosen profession, b) OET test takers find the topics interesting, and as a result, they engage more with test preparation and OET tasks, and c) their familiarity with terminology and content reduces their anxiety while carrying out written and spoken tasks. Similarly, representatives of regulatory bodies appear to value the relevance of OET test to the healthcare context. The representatives stated that *if testing is congruent with practice, that's terrific*; but they pointed out that OET is not expected to test clinical communication skills.

Language as an indicator of workplace readiness

The overseas healthcare professionals who have taken OET are perceived as sufficiently able users of English in their workplace, as there has been no feedback to the contrary.

Some of the quotes provided by the stakeholders in this study demonstrate the usefulness of OET in terms of workplace readiness:

> *As an employee (nurse) in a hospital you are expected to function almost a hundred percent from day one, meaning that you are expected to understand both patients and staff, the latter often speaking very fast and with lots of abbreviations. Preparing for OET helped a lot.*
>
> *It helped me to communicate with patients and work mates effectively and correctly. Because I have gained a lot of good communication styles in a very professional and elegant way.*
>
> *I took patient's history exactly the same way as been taught in OET courses e.g. The patient's pain complain[t] history is the same I used it with my patients ...*
>
> *OET helped me in gaining communication skills with patients and other health professionals. Now I can use some expressions in calming patients, showing empathy to patient which I knew but never used before.*

Table 2 displays health professionals' percentage agreement with the questionnaire statement indicating the ability of OET test takers to use English in a health-related workplace. The percentage agreement column sums up the figures for the 'strongly agree' and 'agree' categories.

Table 2: The ability of OET test takers to use English in a health-related workplace

The employees/colleagues who have taken OET ...	Percentage agreement
... use English effectively in their health-related workplace.	93%
... communicate well with their colleagues.	83%
... understand well what they are told by their colleagues.	83%
... perform well at the writing tasks in their health-related workplace.	83%
... understand well what they read in their health-related workplace.	80%
... communicate well with their patients.	68%
... understand well what they are told by their patients.	65%

After overseas health professionals have been employed, the regulatory boards will receive comments on their language skills. The most common feedback received focuses on the employees' ability to deal with idioms and slang. The board is aware that this is a delicate issue:

> *It's a very fine line ... because there's a ... sense that you need to be able to speak in non-medicalese, but at the same time you need to know when. The level of formality needs to vary. And that must be tricky for a non-native speaker.*

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Conclusion

Looking at data provided by the various stakeholders, several themes emerged with regard to the strengths gained and demonstrated by OET test takers within the workplace. OET test takers are perceived as effective communicators who are able to communicate with stakeholders in the healthcare sector on matters that are both technical and emotional and who can use lay language so that patients can easily understand what they are saying. Another theme which kept emerging from the interviews and open-ended comments on surveys is the need to understand Australian slang, cultural peculiarities, and different accents in order to be able to work better within an Australian context.

These findings provide a useful insight into the appropriacy of OET for healthcare professionals. Further studies will consider candidate performance on OET and similar exams used for the same purpose.

References

Cambridge English Language Assessment (2013) Australian test of medical English to go global, *Cambridge English News*, available online: www.cambridgeenglish.org/news/view/australian-test-of-medical-english-to-go-global/

Greene, J C, Caracelli, V J and Graham, W F (1989) Toward a conceptual framework for mixed-method evaluation designs, *Educational Evaluation and Policy Analysis* 11 (3), 255–274.

The University of Melbourne (2009) *Occupational English Test*, Melbourne: The OET Centre.

APPENDIX 1: Institutes and organisations that recognise OET

Australia

Australian Department of Immigration and Citizenship (DIAC)

OET is accepted by DIAC for various visa categories including skilled migration and student visas.

Boards and Councils

Australian Health Practitioner Regulation Agency (AHPRA)

National agency responsible for the registration and accreditation of 9 of the 12 health professions in Australia. These Boards are:

- Dental Board of Australia
- Medical Board of Australia
- Medical Radiation Practice Board of Australia
- Nursing and Midwifery Board of Australia
- Occupational Therapy Board of Australia
- Optometry Board of Australia
- Pharmacy Board of Australia
- Physiotherapy Board of Australia
- Podiatry Board of Australia

Other regulatory authorities that recognise the OET:

- Australasian Veterinary Boards Council (AVBC)
- Australian and New Zealand Podiatry Accreditation Council (ANZPAC)
- Australian Dental Council
- Australian Institute of Radiography (AIR)
- Australian Nursing and Midwifery Accreditation Council
- Australian Pharmacy Council
- Australian Physiotherapy Council (APC)
- Australian Veterinary Boards Council (AVBC)
- Dieticians Association of Australia (DAA)
- Occupational Therapy Council (Australia and NZ) (OTC)
- South Australian Medical Education and Training (SA MET) Health Advisory Council (HAC)
- Speech Pathology Australia

Universities, Education and Recruitment

- Australian College of Nursing
- Australian Catholic University
- Curtin University
- C.Y. O'Connor Institute
- Deakin University
- Monash University
- Perth Institute of Business and Technology (PIBT)
- Southern Cross University
- Tafe NSW
- The University of Queensland
- University of Newcastle
- University of Notre Dame Australia
- University of South Australia
- University of the Sunshine Coast
- University of Western Sydney
- Geneva Health
- Latitudes Group International
- Recruit-A-Doc

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

New Zealand

- Australian and New Zealand Podiatry Accreditation Council (ANZPAC)
- Dental Council of New Zealand
- Nursing Council of New Zealand
- Occupational Therapy Council (Australia and NZ) (OTC)
- Pharmacy Council of New Zealand
- Physiotherapy Board of New Zealand
- Veterinary Council of New Zealand

Singapore

- Allied Health Professions Council
- Singapore Medical Council

APPENDIX 2: OET administration venues

Africa

Egypt
Cairo

South Africa
Pretoria

Asia

India
Bangalore
Chennai
Kolkata
Mumbai
New Delhi

Japan
Tokyo

Pakistan
Islamabad
Lahore

Europe

Germany
Berlin
Oldenburg

Italy
Catania
Pisa
Telese Terme
Trieste

UK
London
Melton Mowbray

Ukraine
Kharkiv

Middle East

Iran
Tehran

Saudi Arabia
Eastern province

UAE
Dubai

North America

Canada
Vancouver

USA
New York

Oceania

Australia
Adelaide
Brisbane
Darwin
Melbourne
Perth
Sydney

New Zealand
Auckland
Christchurch
Hamilton
Palmerston North

South America

Argentina
Cordoba

Chile
Vina Del Mar

South East Asia

Philippines
Manila

Hong Kong
Singapore

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Book announcement: Validity in Educational and Psychological Assessment

PAUL NEWTON INSTITUTE OF EDUCATION, UNIVERSITY OF LONDON
STUART SHAW CAMBRIDGE INTERNATIONAL EXAMINATIONS

Introduction

For almost 100 years, divergent views on the concept of validity have proliferated. Even today, the meaning of validity is heavily contested. Despite a century of accumulated scholarship, new definitions of validity continue to be proposed, and new 'types' of validity continue to be invented (see Newton and Shaw 2013). Yet, against the backdrop of an evolving measurement and testing landscape and the increased use of assessments across scientific, social, psychological and educational settings, validity has remained 'the paramount concept in the field of testing' (Fast and Hebbler 2004:i).

Validity is universally regarded as the hallmark of quality for educational and psychological measurement. But what does quality mean in this context? And to what exactly does the concept of validity actually apply? What does it mean to claim validity? And how can a claim to validity be substantiated? In a book entitled *Validity in Educational and Psychological Assessment*, which is due to be published by SAGE in early 2014, we explore answers to these fundamental questions.

Validity in Educational and Psychological Assessment adopts an historical perspective, providing a narrative through which to understand the evolution of validity theory from the 19th century to the present day. We describe the history of validity in five broad phases, mapped to the periods between:

1. The mid-1800s and 1920: gestation
2. 1921 and 1951: crystallisation
3. 1952 and 1974: fragmentation
4. 1975 and 1999: (re)unification
5. 2000 and 2012: deconstruction

We explain how each of these phases can be characterised by different answers to the question at the heart of any validation exercise: how much and what kind of evidence and analysis is required to substantiate a claim of validity?

The book comprises six chapters. In Chapter 1 we set the scene for the historical account which follows. Chapters 2 through 5 offer readers a chronological account that delineates the phases of development of validity theory and validation practice. In Chapter 6 we propose a framework for the evaluation of testing policy, which we based on the original progressive matrix from Messick (1980).

Chapter 1: Validity and validation

In Chapter 1 we begin by exploring a range of everyday and technical meanings of validity in order to set the scene for the historical account which follows. This is an account of validity as a technical term of educational and psychological measurement, which is important to bear in mind because the term 'validity' has very many different meanings, some of which are entirely independent of measurement. The main chapters of the book attempt to demonstrate how, even within this relatively narrow conceptualisation, its meaning is still nevertheless contested and resistant to precise definition. Yet it needs to be appreciated, from the outset, that it does mean something quite distinctive in this particular context, even if that 'something' might be difficult to articulate.

Following a discussion of the conventions used in the textbook we present an outline of the history of validity. The historical account is our attempt to describe and to explain how conceptions of validity and validation have evolved within the field of educational and psychological measurement.

Our historical account tends to focus more on concepts of validity theory than on the practice of validation. Good validation practice is the application of good validity theory. In the absence of validity theory there is nothing to guide or to defend validation practice. It is theory that constitutes the rational basis for validation practice. As we discuss each new contribution to the theory of validity, their implications in terms of a positive, operational impact upon validation practice become increasingly apparent.

Chapter 2: The genesis of validity (mid-1800s to 1951)

Chapter 2 covers the first two phases outlined above: a gestational period, from the mid-1800s to 1920; and a period of crystallisation, from 1921 to 1951. The chapter is heavily skewed towards the latter, as the period during which the concept of validity developed an explicit identity or, perhaps more correctly, a range of different identities.

In Chapter 2, we explore early conceptions of validity and validation, focusing particularly upon achievement tests, general intelligence tests, and special aptitude tests. We argue that the emergence of validity as a formal concept of educational and psychological measurement can only be understood in the context of major developments in testing for educational, clinical, occupational and experimental purposes which occurred during the second half of the 19th century and the early decades of the 20th century, most notably in England, Germany, France and the USA. Upon this foundation was proposed the 'classic' definition of validity: the degree to which a test measures what it is supposed to measure.

Although there are numerous accounts of the history of validity theory and validation practice during the early

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

years (e.g. Anastasi 1950, Geisinger 1992, Kane 2001, Shepard 1993) the impression given is often of a period almost exclusively dominated by prediction, the empirical approach to validation, and the validity coefficient. Reflecting on this period, Cronbach (1971) observed that the theory of prediction was very nearly the whole of validity theory until about 1950; a characterisation later endorsed by Brennan (Ed) (2006). Kane (2001) characterised the early years as the 'criterion' phase, where the criterion was typically understood as the thing that was to be predicted.

The impression given by a number of notable chroniclers (e.g. Moss, Girard and Haniford 2006) is that the key developments in validity theory can be traced either to successive editions of *Educational Measurement* (beginning with Lindquist (Ed) 1951) or to successive editions of professional standards documents (beginning with American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education (1954)). We argue that there is a far more interesting story to be told about the early years. We contend that many of the developments in validity theory and validation practice, from the middle of the 20th century onwards, are simply elaborations of earlier insights. The earliest definition of validity was far more sophisticated than the idea of a validity coefficient might suggest, and the earliest approaches to validation were far more complex and involved. Education took a lead in formally defining the concept, and achievement testers, aptitude testers, intelligence testers and personality testers played their role in refining it and developing new techniques for investigating it.

The more interesting story of validity during the early years is one of sophistication and diversity; at least in terms of ideas, if not always in terms of practice. Because of its diversity, though, it is hard to characterise the period succinctly.

Chapter 3: The fragmentation of validity: 1952 to 1974

The diversity of ideas on validity and validation during the early years presented a challenge to test developers and publishers. Given a variety of approaches to validation to choose from, and with even the experts valuing those approaches quite differently, how were professionals in the field to decide what information on test quality they needed to make available to consumers? And, in the absence of agreement upon principles of best practice and specific guidelines about criteria for the evaluation of tests and testing practices, how were test developers and publishers to be held to account?

The first edition of what was to become known as the *Standards* (American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education 1954) was written to make sense of the landscape of the early years. As a consensus statement of the professions, the *Standards* included both implicit standards for thinking about validity and explicit standards for conducting and reporting validation research.

The *Standards* emphasised 'types' of validity, specialised to the contexts of test use: content validity, predictive validity, concurrent validity, and construct validity. If, for example, you needed to validate an interpretation drawn in terms of achievement, then you needed to adopt a particular approach to validation: content validation, which meant establishing a particular kind of validity: content validity. Although these were explicitly described as 'four types of validity' (American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education 1954:13) the *Standards* was a little confused over the matter, also describing them as 'aspects' of a broader conception.

Between 1954 and 1974, the *Standards* was revised twice, in order to respond to constructive criticism, to take account of progress in the science and practice of educational and psychological measurement, and to respond to societal change. Yet, mixed messages continued to be promulgated over the nature of validity. For many who were influenced by the *Standards* during this time, they came to embody and to cement a fragmented view of validity and validation, whereby different uses to which test scores were to be put implied different approaches to validation and even different kinds of validity.

Chapter 4: The (re)unification of validity: 1975–1999

Messick's account of validity and validation became the zeitgeist of late 20th century thinking on validity during the 1980s and 1990s. Developing ideas from Harold Gulliksen and Jane Loevinger, and with the support of allies including Robert Guion, he brought the majority of measurement professionals of his generation around to the viewpoint that all validity ought to be understood as construct validity. His thesis was that measurement ought to be understood (once more) as the foundation for all validity; and therefore that construct validation – scientific inquiry into score meaning – ought to be understood as the foundation for all validation.

Through an extended discussion of Messick's contribution to validity theory, we describe this period in terms of his *triumph* and his *tribulation*. Messick was enormously successful in promoting validity as a unitary concept, in contrast to earlier fragmented accounts. His triumph, therefore, concerned the *science* of validity: he convinced the educational and psychological measurement communities that measurement-based decision-making procedures (i.e. tests) needed to be evaluated holistically, on the basis of a scientific evaluation into score meaning. Enormously problematic, though, was his attempt to integrate values and consequences within validity theory through his famous (if not infamous) progressive matrix. Unfortunately, not only was his account confusing, it also seemed a little confused. His tribulation, it seems fair to conclude, concerned the *ethics* of validity. Messick failed to provide a convincing account of how ethical and scientific evaluation could straightforwardly be integrated.

In retrospect, it seems hard to disagree with the conclusion, drawn by Shepard (1997), that Messick's progressive matrix was a mistake. Having said that, we believe that its underlying intention was an excellent one. It was an attempt to

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

emphasise that the following two questions were both crucial to any thorough evaluation and were inherently interrelated:

1. Is the test any good as a measure of the characteristic it purports to assess?
2. Should the test be used for its present purpose?

Messick's progressive matrix was supposed to explain the relationship between these two questions, and their relation to the concept of validity, but it was muddled. As Messick helped readers to find their way through the ambiguity of the matrix, his presentation became clearer, but also narrower, as scientific questions of test score meaning began to gain prominence while ethical questions of test score use were nudged into the wings.

Unfortunately, Messick's tribulation led to one of the most notorious debates of all time concerning the scope of validity theory. The field is now genuinely split as to whether, and if so how, evidence from consequences ought to be considered part of validity theory – an issue we tackle in Chapter 5.

Chapter 5: The deconstruction of validity (2000 to 2012)

During the 1990s, work on validity and validation was heavily influenced by Messick. The fifth edition of the *Standards* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education 1999) was essentially a consensus interpretation of his position, that is, a unified conception of validity. The *Standards* reflected the prevailing view of the time, a construct-centred approach to validity. Yet, with the turn of the millennium, cracks began to emerge. On one hand, it was unclear how to translate construct validity theory into validation practice. On the other hand, it was unclear whether construct validity was actually the best way to unify validity theory. It seemed that an element of deconstruction might be in order, reflecting the desire to simplify validation practice as well as the desire to simplify validity theory.

In terms of validation practice, this period was characterised by growing consensus over the value of a new methodology for guiding, and simplifying, validation practice. Argumentation, it now seemed, held the key. Kane had developed a methodology to support validation practice, grounded in argumentation (e.g. Kane 1992). This provided a framework, or scaffold, for constructing and defending validity claims. Thus, while Messick defined the claim to validity in terms of an overall evaluative judgement, Kane explained exactly how that claim to validity could be constructed and defended. The argument-based approach took a long time to take root, though, and only began to have a significant impact well into the new millennium. In fact, even having begun to take root, it still proved surprisingly challenging to implement. Goldstein and Behuniak (2011) noted that very few examples are available to the research community of validity arguments for large-scale educational assessments.

In terms of validity theory, this period was characterised by growing controversy, embodied in two major debates. The first concerned the nature and significance of construct validity: a debate over the relatively narrow, scientific issue of score meaning. A critical question was whether construct validity ought to be considered the foundation of all validity, as Messick had argued. Related questions concerned whether all validation needed to be understood in terms of constructs; whether the nomological networks of Cronbach and Meehl (1955) were useful or even relevant to validation; whether validity was a concept more like truth or more like justified belief; whether validity ought to be theorised in terms of measurement; and whether the concept of validity could be applied in the absence of standardised procedures.

The second concerned the scope of validity: a debate over whether the concept ought to be expanded beyond the relatively narrow, scientific issue of score meaning, to embrace broader ethical issues concerning the consequences of testing. Various 'camps' developed: from liberals, who extended the use of 'validity' to embrace social considerations of test score use; to conservatives, who restricted the use of 'validity' to technical considerations of test score meaning.

Chapter 6: 21st century evaluation

The concept of validity has assumed a pivotal role across decades of debate on the characteristics of quality in educational and psychological measurement. Despite this, it has proved extremely resistant to definition. In Chapter 6, we respond to the concerns of the more conservatively minded, who object that the concept of validity is becoming so large as to present an obstacle to validation practice. We do so by proposing a new framework for the evaluation of testing policy. In fact, we see this as a revision of the original progressive matrix from Messick (1980), which we have redesigned to dispel some of the confusion engendered by its original presentation. After first defending the new framework we then provide a more detailed analysis of technical and social evaluation, before considering evaluation within each of the cells respectively.

Validity in Educational and Psychological Assessment is a SAGE publication and will be available from early 2014. The authors believe that this book will be of interest to anyone with a professional or academic interest in evaluating the quality of educational or psychological assessments, measurements and diagnoses.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999) *Standards for Educational and Psychological Testing*, Washington, DC: American Educational Research Association.

American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education (1954) Technical recommendations for psychological tests and diagnostic techniques, *Psychological Bulletin* 51 (2), supplement.

Anastasi, A (1950) The concept of validity in the interpretation of test scores, *Educational and Psychological Measurement* 10 (1), 67–78.

Brennan, R L (Ed) (2006) *Educational Measurement*, 4th edition, Westport: Praeger Publishers.

Cronbach, L J (1971) Test validation, in Thorndike, R L (Ed) *Educational Measurement*, 2nd edition, Washington, DC: American Council on Education, 443–507.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

Cronbach, L J and Meehl, P E (1955) Construct validity in psychological tests, *Psychological Bulletin* 52 (4), 281–302.

Fast, E F and Hebbler, S (2004) *A Framework for Examining Validity in State Accountability Systems*, Washington, DC: Council of Chief State School Officers.

Geisinger, K F (1992) The metamorphosis of test validation, *Educational Psychologist* 27 (2), 197–222.

Goldstein, J and Behuniak, P (2011) Assumptions in alternate assessment: An argument-based approach to validation, *Assessment for Effective Intervention* 36 (3), 179–191.

Kane, M T (1992) An argument-based approach to validity, *Psychological Bulletin* 112 (3), 527–535.

Kane, M T (2001) Current concerns in validity theory, *Journal of Educational Measurement* 38 (4), 319–342.

Lindquist, E F (Ed) (1951) *Educational Measurement*, Washington, DC: American Council on Education.

Messick, S (1980) Test validity and the ethics of assessment, *American Psychologist* 35, 1,012–1,027.

Moss, P A, Girard, B J and Haniford, L C (2006) Validity in educational assessment, *Review of Research in Education* 30 (1), 109–162.

Newton, P E and Shaw, S (2013) Standards for talking and thinking about validity, *Psychological Methods* 18 (3), 301–319.

Shepard, L A (1993) Evaluating test validity, *Review of Research in Education* 19, 405–450.

Shepard, L A (1997) The centrality of test use and consequences for test validity, *Educational Measurement: Issues and Practice* 16 (2), 5–8.

Caroline Clapham IELTS Masters Award 2013

For over a decade now, the IELTS partners have presented the Caroline Clapham IELTS Masters Award annually to the Master's level dissertation or thesis in English which makes the most significant contribution to the field of language testing.

Recently, the IELTS Research Committee announced the selection of Benjamin Kremmel as the winner of the 2013 award. His dissertation investigated the factors that predict second language reading comprehension, and also contributed to the literature on the replicability of research findings. One reviewer sums up the merits of the dissertation, describing it as 'an ambitious, sophisticated and very thought-provoking dissertation which offers well-grounded evidence for the conclusions that are reached, and which is indeed an improvement, as claimed, on the reference study which the present study aims to replicate and improve… [c]arried out with exemplary thoroughness and reported in a lucid, considered way.' The dissertation was supervised by Tineke Brunfaut, and was submitted to Lancaster University. The abstract for the dissertation appears below.

Benjamin will be presented with his award – a certificate and a cheque for £1,000 – at the 2014 Language Testing Research Colloquium in Amsterdam, The Netherlands. Qualified individuals who would like to join the 2014 competition are invited to visit: ielts.org/researchers/grants_and_awards/ielts_masters_award.aspx for details of the competition and submission guidelines.

Explaining Variance in Reading Test Performance through Linguistic Knowledge: The Relative Significance of Vocabulary, Syntactic and Phraseological Knowledge in Predicting Second Language Reading Comprehension

Benjamin Kremmel

Despite the acknowledgement of the importance of second language (L2) knowledge for reading ability in applied linguistics, to date there has been little agreement among researchers as to which aspect of this knowledge is most crucial for L2 reading comprehension. The present dissertation investigates the relative significance of the components vocabulary knowledge and syntactic knowledge in explaining variance in reading test performances. Examining closely and partly replicating a case study by Shiotsu and Weir (2007), it aims to examine which of the two components is a better predictor of reading ability. It thereby problematises traditional theoretical conceptualisations and practical operationalisations of vocabulary and structural knowledge and questions whether a dichotomous distinction between the two and thus any judgement as to which of the two is a more important contributor to L2 reading is legitimate and indeed feasible.

The dissertation thus presents a study which analyzes the test scores of 418 Austrian learners of English on a syntactic knowledge test, a vocabulary knowledge test, a test of multi-word expressions and a reading test by means of multiple regression models and structural equation modeling. In so doing, the study examines whether a broader construct definition of linguistic knowledge taking phrasal expressions into account provides a more comprehensive construct representation and valuable insights into the role of phraseological knowledge in reading comprehension.

The study shows that the syntax measure used in the case study is problematic and its findings concerning the superiority of syntactic knowledge in predicting reading test performance could not be corroborated. However, the study attempts to demonstrate that the notion of 'lexicogrammar' (Sinclair 2004), viewing lexical and syntactic knowledge as a cline rather than a dichotomy, should be adopted in the field of language testing and that it might be best to refrain from simplistic claims about the prevalence of one of the traditional components over another.

© UCLES 2013 – The contents of this publication may not be reproduced without the written permission of the copyright holder.

© UCLES 2013 CE/2089/3Y11

541708892 6

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

Contents:

Editorial	2
Delta trainers' cognitions and pedagogical practices Simon Borg	3
Washback effect of Cambridge English: Key for Schools in São Paulo, Brazil: A case study Matilde Virgínia Ricardi Scaramucci and Eliana Kobayashi	13
Young learner perceptions of test taking in Japan Jennifer M Toews-Shimizu	21
Stakeholders' perceptions of the Occupational English Test (OET): An exploratory study Ivana Vidaković and Hanan Khalifa	29
Book announcement: Validity in Educational and Psychological Assessment Paul Newton and Stuart Shaw	33
Caroline Clapham IELTS Masters Award 2013	36

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

Cambridge English
Language Assessment
1 Hills Road
Cambridge
CB1 2EU
United Kingdom
Tel. +44 1223 553997

ALTE Association of Language Testers in Europe

A DIVISION OF CAMBRIDGE ASSESSMENT

© UCLES 2013 – this publication may not be reproduced without the written permission of the copyright holder