ETiC

ENGLISH TEACHING IN CHINA
在华英语教学

Issue 7 • April 2016

RESEARCH • REFLECTIONS • REVIEWS

A peer-reviewed journal for teaching professionals

CONTENTS

Welcome to this assessment-themed 7th issue of English Teaching in China (ETiC), and to our first as Editors. We have had the good fortune of inheriting, from our predecessors Amanda and Mark, a **journal in rude health. It is a testament to their stewardship that we** have made very few changes to how the journal is run. Thank you!

This is, however, an opportunity to take stock, and to re-examine our identity as a journal. ETiC grew out of the Language Centre at Xi'an Jiaotong-Liverpool University (XJTLU): Issue 1 was written and produced solely by teachers at XJTLU. Now, it would be fair to say that the emphasis has shifted: ETiC , with its Editorial Board of leaders in the field, with peer-reviewers across 4 continents, with authors from diverse teaching contexts, has grown out of XJTLU. But throughout this transition, one constant has remained: the authors who publish in this journal are the readers who read it. They are, overwhelmingly, practitioners in the field who want to share what they are doing with others in similar situations. The issues our authors grapple with in the following pages are the very same ones our readers grapple with on a daily basis. As Editors, we see ETiC's role primarily as helping those who want to share what they are doing, share what they are doing, regardless of previous publishing experience. The current issue, we believe, epitomises this approach.

We said earlier that we have made very few changes to how ETiC is run. The minor changes that we have made, however, have been made to aid this process of getting into print the ideas and practices of those who may not have published before. So, this will be the last issue with a theme: from now on, we will be considering all contributions relevant to all aspects of English teaching in China. In addition, there will be no submission deadlines as such: submissions, when received, will be considered for the next workable issue, starting with Issue 8 later this year.

But for now, over to the authors of Issue 7. Enjoy!

Jonathan and Samantha

Jonathan.Culbert@xjtlu.edu.cn
Samantha.Ng@xjtlu.edu.cn

CREDITS

CO-EDITORS
Jonathan Culbert
Samantha Ng

ASSOCIATE EDITORS
Debra Jones
Layla Shelmerdine
Richard Galletly
Seth Hartigan
Jackie Hemingway

CHIEF COPY-EDITOR
Jennifer Howard

COPY-EDITORS
Ann Brantingham
Micah Park
Nick Merwitzer
Paul Smit
Rick Foster

PEER REVIEWERS
Bruce Applebaum
David Fryatt
Diane Du
Eoin Jordan
Helen Gasking
Jon Ford
Micah Park
Michelle Ives
Paul Meier
Ralph Hughes
Simon Smith

TRANSLATORS
Junshuan Liu
Shu Deng (Jackie)

CIRCULATIONS MANAGER AND WEBMASTER
Rainie (Minghao) Zhang

EDITORIAL BOARD
Andy Kirkpatrick
(Griffith University)
Stuart Perrin
(Xi'an Jiaotong-Liverpool University)
Don Snow
(Duke Kunshan University)
Leslie Gryce Sturino
(Xi'an Jiaotong-Liverpool University)
Labao Wang
(Soochow University)
Amanda Hilmarsson-Dunn
(Former Editor)
Mark Critchley
(Former Editor)

AUTHORS

DOUGLAS GLENN
Douglas Glenn was the Secondary Librarian at Suzhou Singapore International School from 2011 to 2015. He is currently a consultant living in Panama. his main professional interest is helping secondary students master academic research.

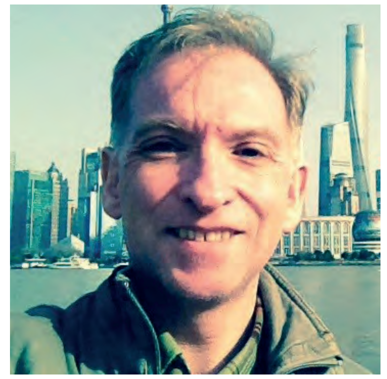
dglenn047@gmail.com

JACKIE HEMINGWAY
Jackie Hemingway has taught EAP at XJTLU since September 2011. Prior to joining XJTLU she taught General and Business English in several countries including Austria, Spain, Japan and Oman.

jacqueline.hemingway@xjtlu.edu.cn

SIMON DAWSON
Currently EAP Tutor at University of Nottingham in Ningbo; soon to be EAP tutor at Bristol University. I teach EAP courses and design and develop EAP assessments. Recent assessment projects include the revision of rating scales for seminar discussion skills and interview speaking skills.

Simon.dawson@nottingham.edu.cn



NIGEL DIXON
Nigel Dixon is an EAP Tutor and former Examinations Officer at XJTLU. He has worked on test preparation and taught English in China, Indonesia, Turkey, Thailand and the UK, and has recently gained experience as a language test candidate in Chinese HSK exams.

nigel.dixon@xjtlu.edu.cn

SETH HARTIGAN
Seth Hartigan has taught Academic English for the last ten years at XJTLU, Renmin and Tsinghua Universities. His academic interests include the philosophy of education and the science of learning.

seth.hartigan@xjtlu.edu.cn

GRAHAM MORTON
Graham Morton currently teaches English Language Acquisition at SSIS. He has an MSc in TESOL and has taught at international schools inTaiwan, Vietnam and China. His professional interests include vocabulary development, and contrastive rhetoric for international school students.

grahammorton@ssis-suzhou.net

SARAH BUTLER
Sarah Butler is Associate Director of Pre-sessional English courses at CU Services Coventry University. Her areas of interest include teacher development and learner autonomy.

ab6631@coventry.ac.uk

ROSS THORBURN
Ross Thorburn has worked in the language teaching industry in China for the past nine years. He currently works as a Training & Development Manager, decreasing staff turnover through providing high-quality training and internationally recognized qualifications to both teachers and managers.

ross.thorburn@yahoo.com

DEBRA JONES
Debra Jones taught English and Academic Skills in Japan and China for 12 years before taking up her current post as In-sessional Course Leader in the Academic Skills Centre at Bath University. Her research interests are in feedback and autonomous learning.

daje17@yahoo.co.uk

ASSESSMENT AS LEARNING: THE NEXT STEP IN FORMATIVE ASSESSMENT

Debra Jones

ABSTRACT
The value of formative assessment or Assessment for learning (AfL) in improving student performance is widely accepted and practised in L2 classrooms. However, its potential for developing learner autonomy is not always realised. This article presents an overview of Assessment as Learning (AaL), a recent development which promotes autonomous learning by shifting responsibility from teachers to students. The principles of AaL and how it works in practice will be introduced, followed by a discussion of some benefits and potential problems in implementing it. The article concludes that AaL should be incorporated into both formative and summative assessment processes and presents some ideas for achieving this.

摘要
形成性评估或“以评促学”（AfL）的方法旨在提高学生的学习表现，其价值已为第二语言课堂广泛接受和采用。然而，这些方法对提高学习者自主学习能力的潜在作用并非总是得以发挥。本文概述了一种新的方法 –“评估作为学习”（AaL）。该方法强调通过将学习责任从教师转到学生来促进学生自主学习能力的提高。文章首先介绍了AaL的原则及其在实践中的应用，然后对其益处及实际应用中潜在的问题进行了讨论，最后指出AaL应与形成性评估和总结性评估相结合，并对如何结合提出了一些设想。

Assessment in English language teaching, as in other fields, has traditionally focused on summative evaluation of a finished product, or Assessment of Learning (AoL). Since the 1990s, following the work of Black and Wiliam (1998), there has been increasing interest in formative assessment, or Assessment for Learning (AfL), where feedback is given during the process of learning to help students improve their performance and provide diagnostic information for teachers (Lee & Coniam, 2013). Arguably, however, AfL is still teacher-centred. Even where an element of peer review is included, students are primarily responding to comments and advice from their teachers on how to improve. As Earl and Katz (2006) point out:

If all feedback does is provide direction for what students need to do – that is, if the feedback doesn't refer to students own role in moving forward to the next stage of learning – they will be perpetually asking questions like 'is this right?' 'Is this what you want?' (p. 48)

In other words, although formative feedback improves performance, it does not necessarily develop learner autonomy. Learners are not making decisions for themselves; they are simply following the direction of the teacher. More importantly, formative feedback tends to be product focused rather than contributing to the process of learning since feedback is primarily used to improve a piece of work with the aim of achieving a higher grade. Although teachers hope that formative feedback will be carried forward to future assignments, studies have shown that students believe the purpose of formative feedback is to improve the finished product rather than enhance future learning (Duncan, 2007; Wallis, Jones, & Xu, 2014).

More recently, the idea of Assessment as Learning (AaL), has emerged as a sub-category of formative assessment (Volante & Beckett, 2011, p. 247), exploiting the potential of formative feedback to develop autonomy by transferring responsibility for learning from teacher to student (Lafave, Katz, & Vaughan, 2013). AaL is student-focused, "emphasizing assessment as a process of metacognition (knowledge of one's own thought processes) for students" (Earl & Katz, 2006, p. 41). AaL enhances metacognitive skills by making students think about what they are learning and how successful they have been through reflection, monitoring and self-assessment activities (Lee, 2014). As such, AaL adopts a constructivist approach, whereby learning is an active process of constructing meaning rather than a transfer of knowledge from teacher to student (Earl & Katz, 2006).

This article presents an overview of AaL, outlines how it can be implemented in the classroom and discusses its benefits and possible problems, with particular reference to the Chinese context.

WHAT DOES IT MEAN IN PRACTICE?

Assessment as Learning acknowledges the role of students in the assessment process (Lee & Coniam, 2013, p. 35). Students engage in a variety of tasks to monitor their own progress and make adjustments based on their understanding of their own learning needs. In other words, students "learn to learn" through active involvement in their own assessment (Earl & Katz, 2006).Three stages can be identified:

(1) Planning or goal setting. This can take the form of independent learning plans (ILPs), where students identify their strengths and weaknesses at the start of the course, set goals for addressing their perceived needs, and propose steps they will take to achieve their goals. It could also involve setting goals for future improvement based on peer or teacher feedback on an assignment, a process generally referred to as "feeding-forward", where students engage with and act on feedback in future assignments (Duncan, 2007; Jones, 2011).

(2) Regulation or reflection, which involves monitoring progress. After setting goals, learners regularly evaluate their progress through continuous review of their ILPs or by assessing how successful they have been at feeding-forward. Reflective journals, either handwritten or online, or e-portfolios also provide an opportunity for monitoring learning by allowing students to look back at previous work and reflect on their progress (see Kim, 2013; Yastibas & Yastibas, 2015). These monitoring activities should be built into the course and, ideally, done during class time to allow teachers to check that students understand the concepts of reflection and goal-setting. Once students cultivate the habit of monitoring their learning, it can form part of their independent learning activities outside class.

(3) Evaluation, where students themselves participate in evaluating the product of their learning. This allows for the possibility of student involvement in the summative assessment process. Teachers can facilitate this by providing models or checklists for students to assess their performance. Checklists can be based on the course learning outcomes or task descriptors. However, Brindley (1984, as cited in Nunan, 2013, p. 57) has argued that adult learners should have a say in how they are assessed. Thus, having acquired the skill of self-evaluation, students should be encouraged to assess their work using their own assessment criteria based on their individual needs.

The ultimate goal of AaL is developing independent learners capable of assessing their own strengths and weaknesses and taking responsibility for their future learning. However, guidance from teachers is needed, at least in the initial stages, since students are often unfamiliar with reflective tasks, particularly those used to a more passive-learning, ▷

teacher-centred education (Lucenta, 2011; Kim, 2013). The teacher's role is to help students acquire the skills they need to perform these reflective tasks and develop the habit of self-monitoring and assessment. This can be done by creating opportunities for students to practise the skills and by giving feedback. Unlike traditional formative feedback, in AaL feedback should focus not simply on what students have produced and how they can improve it, but on their degree of understanding of why they have done it and what they have learned from it (Kim, 2013).

BENEFITS OF AN AAL APPROACH

DEVELOPING HIGHER-ORDER COGNITIVE SKILLS

Evaluation and critical reflection are identified as key higher-order thinking skills according to Bloom's taxonomy (Ramsden, 2003, as cited in Kim, 2013) and necessary requirements in the development of a skilled workforce and informed citizenry in the rapidly changing world of the 21st century (Wall, 2015). Despite the stated aims of the Guidelines for Basic Education Curriculum Reform, issued in 2001, which emphasize "cultivating higher-order skills and broadening students' learning experience" (Zhong, 2005 cited in Yan, 2015, p. 6), top-down, teacher-centred pedagogical practices have remained unchanged in most Chinese high schools, largely as a result of negative washback from the university entrance exam or Gaokao (Yan, 2015). Consequently, Chinese students often graduate high school without acquiring, or even being aware of, these higher-order thinking skills. AaL can foster metacognitive awareness, which helps students develop the skills of evaluation and reflection they will need in the future.

ENHANCING MOTIVATION FOR LEARNING

In addition, due to the intensive experience of studying for the Gaokao, Chinese students often emerge from high school with a lack of interest in learning (Lucenta, 2011). Studies suggest that reflection and self-assessment activities can be motivating for students (Finch & Taeduck, 1998; Yastibas & Yastibas, 2015) as they personalize and individualize learning and give learners a measure of control. Motivated students are more likely to develop a positive attitude to learning and embrace the life-long learning ethos necessary in the "rapidly changing, complex, information overloaded world" (Wall, 2015, p. 233).

TRANSFERRING SKILLS AND KNOWLEDGE

Adapting knowledge to other contexts has been seen as one of the main goals of education (Perkins et al., 2000, as cited in Allan & Driscoll, 2014), but transferring knowledge and skills from one course to another is something students seem to find difficult, to the despair of their teachers. The experience of Chinese high school students is dominated by the Gaokao, which encourages teaching to the test and over-emphasis on passing exams in order to progress to the next stage of their academic careers (Lucenta, 2011; Lee & Coniam, 2013; Yan, 2015). When these students progress to university, they are likely to retain this focus on final assessment, so achieving a high grade in a coursework essay takes priority over the learning that takes place during the process of writing the essay. In fact, as a result of their previous educational experience, "learning" is often equated with a high grade (Lee & Coniam, 2013, p. 36). Although this issue of transferability is not unique to Chinese students, the exam-driven education system accentuates the problem (Yan, 2015), making it hard for students to understand the need to transfer skills. After all, if skills learned in one class or during one semester are perceived to be useful only in terms of passing the final summative assessment, then students will not see the need to retain them to facilitate future learning.

Reflecting on learning, one of the cornerstones of AaL, aids skills transfer by helping students think about past, present and future learning, and thus understand learning as a continuum (Allan & Driscoll, 2014). If students understand the process of learning and focus on the skills they are acquiring during the process, they will be better equipped to transfer these skills to future courses. In this way, AaL promotes skills transfer by focussing students on the process, as opposed to the product-centred approach inherent in conventional assessment methods (Earl & Katz, 2006; Yancey, 1998, as cited in Allan & Driscoll, 2014).

SOME PROBLEMS

For teachers, there is often a tension between formative and summative assessment (Harlen, 2005, as cited in Volante & Beckett, 2011; Lee & Coniam, 2013). Teachers are under pressure to be accountable to stakeholders through standardized testing, particularly in the high-stakes context of higher education (Cross & O' Loughlin, 2013). Students and parents tend to be grade-focused, valuing end results over formative assessment, while institutions are often valued and financed based on their results rather than on how much learning has taken place. Despite this, there has been a move away from a "teach, test and hope for the best" approach (Volante & Beckett, 2011, p. 240), and formative assessment techniques such as feedback on first drafts of essays are now common in L2 writing classes. The importance of formative assessment has also been acknowledged at the institutional level and, as Cross and O' Loughlin (2013) point out, many universities have policies requiring students to be given formative feedback on assignments (see University of Liverpool, 2014). Assessment as learning strategies such as self-assessment checklists, ILPs and reflective journals are also practised, but usually at the discretion of individual teachers rather than integrated into the curriculum. Even if such activities are incorporated into the syllabus, an emphasis on summative assessment criteria feeds into student focus on grades, and producing a product that matches descriptors in order to pass, rather than on learning (Kohn, 2011, as cited in Glenn & Morton, 2015).

Related to this, Lee & Coniam (2013) have highlighted the problems of formative assessment in an exam-driven culture such as China, where students may not value tasks that are not graded. Learning styles and perceptions of effective teaching and learning are determined by previous education experience (Brindley, 1984, as cited in Finch & Taeduck, 1998), so students unfamiliar with reflective or self-assessment tasks may reject an AaL approach (Finch & Taeduck, 1998). It has been reported that Chinese students, for example, seem to struggle with tasks such as reflective journals, failing to understand their value and being uncertain of what they are meant to do (Kim, 2013).

SOME SUGGESTIONS

Students' reluctance to engage with anything not directly being assessed is a common frustration of teachers, as is their seeming reluctance to transfer skills and knowledge to a new context. To change this mentality, students need help to see learning as an ongoing process and to understand what they are learning, how they are learning it and why, rather than perceiving learning as something that ends when the exam is over or the coursework is submitted. Simply telling them is not enough. Incorporating some AaL skills such as reflection into assessment practices would emphasise process as well as product and encourage students to understand the value of these reflective activities.

One suggestion is to ease the tension between often rigid, summative assessment frameworks and the potential of formative assessment to improve learning (Cross & O'Loughlin, 2013, p. 593) by reducing the number of summative assessments. This would allow more time for formative assessment and AaL activities in the classroom and give both teachers and students time to reflect and act on feedback.

Another possibility would be to investigate ways of summatively assessing tasks such as reflective journals (see Kim, 2013). Grading these tasks would ensure students took them seriously and through doing the tasks, develop the habit of self-evaluation and reflection. Eventually, they may come to understand the value of these activities for their own sake, not just as a means of improving their grades. It has been suggested that acquiring the habit of using skills, not simply learning them, is more likely to transfer to a new context (Wall, 2015), so cultivating the habit of self-reflection should be an important learning outcome on any course. One difficulty with this approach would be devising criteria to evaluate self-assessment activities. Kim (2013, p. 258) has emphasized the importance of ipsative feedback, which compares current performance with previous work, emphasizing improvement over time (process) rather than measuring ▷

the gap between performance and the required standard laid down in the descriptors (product). Such criteria would be difficult to develop, as it is relative rather than absolute, and may cause problems with standardisation. While standardised testing is important and necessary in a high-stakes environment such as higher education, there is a danger that testing, based on measuring performance against standardised criteria, is done at the expense of learning. As Costa (1991) puts it, "What was educationally significant and hard to measure has been replaced by what is insignificant and easy to measure", so there is a danger of measuring "how well we have taught what is not worth learning" (p. 38). If higher-order thinking skills such as evaluation and reflection are worth learning, then it must be worthwhile finding a way to assess them, however challenging this may prove to be.

A third suggestion is to change who assesses the product by making assessment a collaborative process involving learners and teachers working together (Volante & Beckett, 2011) rather than something teachers do to students. Including an element of student self-assessment as part of their final grade, in addition to the objective criteria applied by the teacher, would motivate students to engage with reflective activities. One objection here could be the validity and reliability of the assessment if self-evaluation is included. However, it has been suggested that assessments should be considered valid if they "inform subsequent phases of teaching and learning" (Moss, 2003, as cited in Cross & Loughlin, 2013, p. 593). Self-assessment activities certainly qualify as valid by this definition.

CONCLUSION

The aim of this article was to introduce a different approach to assessment, present some of its advantages and problems and suggest ways to include it in current assessment practices. Adopting an AaL approach is undoubtedly challenging for teachers and students alike, due to its unfamiliarity and the need to re-think conventional approaches to assessment. However, the potential benefits of AaL are such that incorporating some of its elements into assessment practices is necessary to ensure students gain the higher-order thinking and life-long learning skills required in the 21st century.

REFERENCES

Allan, E. G. & Driscoll, D. L. (2014). The three-fold benefit of reflective writing: Improving program assessment, student learning and faculty professional development. Assessing Writing, 21, 37-55. doi:10.1016/j.asw.2014.03.001

Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. London: Kings College.

Costa, A. (1991). A vision of restructured schools. In P.A. Barrett (Ed.) Doubts and certainties: Working together to restructure schools (pp. 29-47). Washington, D.C.: National Education Association. Retrieved from: http://files.eric.ed.gov/fulltext/ED336867.pdf

Cross, R. & O'Loughlin, K. (2013). Continuous assessment frameworks within university English Pathway Programs: Realizing formative assessment within high-stakes contexts. Studies in Higher Education, 38(4), 584-594. doi:10.1080/03075079.2011.588694

Duncan, N. (2007). "Feed-forward": Improving students' use of tutors' comments. Assessment and Evaluation in Higher Education, 32(3), 271-283. doi:10.1080/02602930600896498

Earl, L. & Katz, S. (2006). Rethinking classroom assessment with a purpose in mind. Western and Northern Canadian Protocol for Collaboration in Education. Retrieved from: http://www.wncp.ca/english/subjectarea/classassessment.aspx

Finch, A. & Taeduck, H. (1998). Oral testing and self-assessment: The way forward? Andong National University. Retrieved from http://anu.andong.ac.kr/~hyun/now/oral%20testing.html

Glenn, D. & Morton, G. (2016). The use of IB assessment rubrics; Teacher and student views compared. English Teaching in China, 7, 28-33.

Jones, D. (2011). Feedback in academic writing: Using feedback to feed-forward. Language Education in Asia, 2(1), 120-133.

Kim, A.K. (2013). Reflective journal assessment: The application of good feedback practice to facilitating self-directed learning. Journal of Hospitality, Leisure, Sport and Tourism Education, 13, 255-259. doi: 10.1016/j.jhlste.2012.04.004

Lafave, M. R., Katz, L., & Vaughan, N. (2013). Application of 'Earl's assessment as, assessment for and assessment of learning model' with orthopaedic assessment clinical competence'. Athletic Training Education Journal, 8(4), 109-114. doi:10.4085/0804109

Lee, I. (2014). 21st century English language assessment: Towards assessment as learning in L2 classrooms. Paper presented at the 34th Annual Thailand TESOL International Conference, Chiang Mai, Thailand. Retrieved from http://thailandtesol.org/Menu/Conference/2014%20ThaiTESOL%20Conference/data/2014plenary.pdf

Lee, I. & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. Journal of Second Language Writing, 22, 34-50. doi:10.1016/j.jslw.2012.11.003

Lucenta, L. (2011). China's higher education lacks higher learning. The Phi Delta Kappan, 93(4), 6-7.

Nunan, D. (2013). Conceptualizing second language teaching and learning. In D. Nunan (Ed.) Learner-centered English language education: The selected works of David Nunan (pp. 46-81). New York, NY: Routledge.

University of Liverpool. (2014). Code of practice on assessment: Appendix N: Policy on feedback on assessment 2014-15. University of Liverpool. Retrieved from http://www.liv.ac.uk/media/livacuk/tqsd/code-of-practice-on-assessment/appendix_N_cop_assess.pdf

Volante, L. & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. Canadian Journal of Education, 34(2), 239-255.

Wall, T. F. (2015). The transferability of higher order cognitive skills. Procedia – Social and Behavioural Sciences, 174, 233-238. doi:10.1016/j.sbspro.2015.01.652

Wallis, T., Jones, D., & Xu, Y. (2014). Written correction techniques at an English medium university in China: What works best? English Teaching in China, 4, 25-30.

Yan, C. (2015). We can't change much unless the exams change: Teachers' dilemmas in the curriculum reform in China. Improving Schools, 18(1), 5-19. doi:10.1177/1365480214553744

Yasitbas, A. E. & Yastibas, G. C. (2015). The use of e-portfolio-based assessment to develop students' self-regulated learning in English language teaching. Procedia – Social and Behavioural Science, 176, 3-13. doi:10.1016/j.sbspro.2015.01.437

Author Biography

Debra Jones taught English and Academic Skills in Japan and China for 12 years before taking up her current post as In-sessional Course Leader in the Academic Skills Centre at Bath University. Her research interests are in feedback and autonomous learning.

Email:
daje17@yahoo.co.uk

PREPARING LECTURE-STYLE TEXTS FOR UNIVERSITY LISTENING ASSESSMENTS

Nigel Dixon

ABSTRACT
Short lectures are an appropriate task for university listening assessments, but to produce quality tests, a number of issues should be considered at different stages of the test-preparation process. This article discusses the features of spoken language in lectures and outlines options for preparing texts which balance the need for authenticity with suitability for use as test material. Evidence for appropriate delivery speed for lectures is also considered. Finally, there is discussion of the advantages and drawbacks of different tasks that may be used to evaluate students' performance in lecture-based listening tests.

摘要
“（简）短讲座”很适合被用于大学听力测试。然而，要制作高质量的试题，应在考试准备的不同阶段对很多因素予以考虑。本文讨论了此类讲座的口语特点，并概述了在准备听力文本的过程中，如何兼顾此文本真实性与适用性。同时，本文也考虑了一些适合此类讲座的语速的例证。最后，文章讨论了在此类听力测试中，用于衡量学生听力水平的各类题型的利弊。

INTRODUCTION

When preparing listening comprehension assessments in a university context, it makes sense to focus on the tasks that students need to perform in their academic studies. Although students will encounter interactional situations requiring them to listen and respond (Lynch, 2011), it is lectures that remain the main method of university teaching (Lee, 2009, p.42).
It is generally accepted that listening to lectures places heavy demands on students' listening comprehension skills (Thompson, 2003, p.5), and thus, the ability to understand lectures is a key measure of students' second language competence (Flowerdew & Miller, 1992, p.60). Short lectures are therefore a logical choice of text for assessment of students' listening comprehension skills. However, for test writers, there are important considerations at different stages of test design. This paper will discuss issues related to text preparation and assessment tasks.

The Common European Framework of Reference of Languages (CEFR) B2 level is accepted as the minimum at which individuals can cope with university studies, while B1 level users may be admitted to pre-study programmes, such as foundation years prior to undergraduate courses. I will therefore make reference to the descriptors for B1 and B2 (Council of Europe, 2001) when discussing appropriate texts.

TEXT PREPARATION

There are several ways of generating lecture texts for listening exams, which can be placed on a continuum from most to least authentic (Figure 1). Rost (2005) argues in favour of using authentic samples of speech on the grounds of test validity: "To assess learners' listening ability, we need to focus on those aspects of proficiency and comprehension that are unique to listening" (Rost, 2005, p.170). In addition, as Flowerdew & Miller point out, features of spoken language in lectures also facilitate the task of comprehension (1997, p.34). These include "false starts, redundancies and repetitions", and short or incomplete clauses linked by pauses or simple conjunctions such as so and okay (1997, p.33). These features give listeners more time to process information, whilst the simpler grammatical structures are easier for them to parse in real time. By contrast, written articles commonly contain greater lexical density and use of embedded clauses than speech (Nesi, 2001), making them more difficult to understand if spoken, as well as sounding unnatural. This leads to the logical conclusion that reading an unaltered written text aloud as test input is unsatisfactory on the count of comprehensibility as well as authenticity.

On the other hand, the authentic option of using recordings of genuine lectures also has drawbacks. Firstly, such recordings may not be suitable for tests. Lectures publicly available on websites are generally too long to use in entirety within the time constraints of exams, and using short extracts creates unfair difficulties for listeners in orienting to the topic. In addition, public lectures are often extra-curricular events, with no intention that the audience should take notes or recall detailed information. They may also lack sufficient salient main ideas for test writers to use as the focus of comprehension

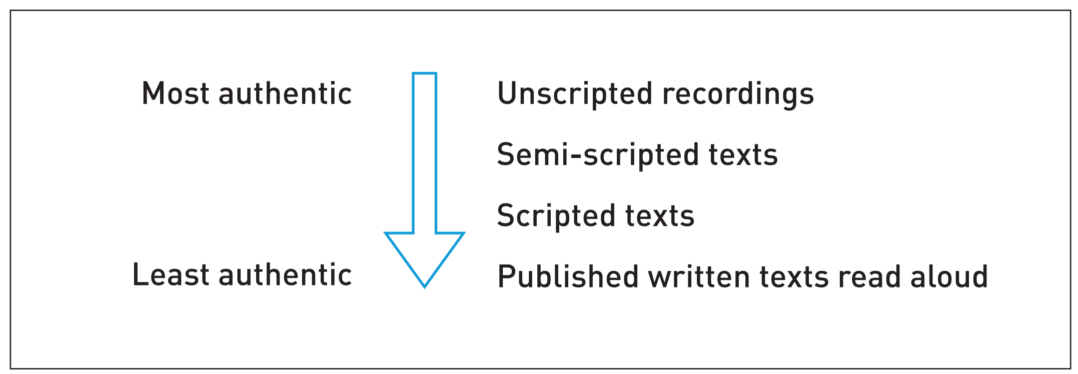


Figure 1: Lecture texts for listening assessments.

questions. However, some of these difficulties might be countered by editing or adding some scripted material to the transcripts of authentic lectures, and then re-recording them.

The third option of preparing scripted texts allows more control over text content, and skilled writers can reproduce at least some of the features of natural speech. This can be done by taking notes on written sources, making an outline, and then writing a final draft in spoken style. The writer can control the complexity of syntax, and may also add features to reduce difficulty, including selected repetition or restatement of main points (Flowerdew & Miller, 1996), and explicit signposting. The latter is necessary to support B2 level learners according to the CEFR descriptors (Council of Europe, 2001, p.8). Nonetheless, in my own experience, writing natural-sounding speech is very challenging, and even scripted materials in EAP textbooks have been criticized for lack of similarity to authentic lecture samples (Flowerdew & Miller, 1997; Thompson, 2003).
A final option is the semi-scripted lecture, in which a speaker improvises from an outline. This has the advantage of giving test-writers control over content while achieving features such as false starts and hesitations which are difficult to 'script'. However, although the technique is fairly well-established, examples in testing guidebooks such as Buck (2001) are limited to simple talks or dialogues on everyday topics rather than lectures, and commercial tests such as IELTS and TOEFL also still use scripted lectures. In a recent account, Clark (2014) describes the use of semi-scripted lectures in a college placement test, in which the volunteer speakers themselves selected topics and sources, and prepared outlines. The resulting lectures fulfilled the aim of creating natural, authentic-sounding speech, but five of the thirteen lectures were deemed unsuitable for the purpose for reasons of length, content, interest or suitability as a listening test (Clark, 2014, pp.9-10). I therefore suggest that given the relatively unproven nature of semi-scripted lectures, test-writers could consider trialling the technique with lower stakes classroom materials before attempting test recordings. For reasons of confidence, the speaker who records a semi-scripted lecture will ideally be experienced in giving presentations. Even then, the outline notes will need to be thorough, as speakers will probably lack the background knowledge to extemporize at length. Notes will likely be in bullet form, using indentation to distinguish key points and supporting examples and details. They should provide key words and phrases, but not complete sentences, in order to maximize the amount of natural spoken syntax. The test preparer may also indicate selected points which speakers should repeat or restate, and specify how much explicit signposting to use for bullet points. The latter two features may not be strictly consistent with natural lecture delivery in L1 settings. Thompson (2003, p.11), for example, found that explicit signposting is far less frequent in authentic lectures than in material typically recorded for EAP learners. Nonetheless, test-writers may consider signposting and some repetition necessary to achieve a degree of difficulty appropriate for L2 users at CEFR B1 or B2 levels. Finally, it may be useful to reassure speakers that natural hesitations are desirable and part of the 'authenticity' of the recording.

Even after taking the steps above, it seems unlikely that a perfectly usable version could be obtained at one attempt. In the article which first introduced the technique of semi-scripting material, Geddes & White therefore recommended recording two complete takes to begin with, and editing them together into the final product (1978, cited in Carr, 2011, p.86).

SPEED AND LENGTH

According to CEFR descriptors, students should be able to follow "standard spoken language" at B2 level, and "clearly articulated" speech at B1 level (Council of Europe, 2001, p.8). In other words, the listening input for the assessment should be at normal or near normal speed. In practice, this is not easy to define. To begin with, speaking speed may vary according to speech act and context. Table 1 compares rates of speech in four sets of data involving lectures and other recorded monologues, measured in words per minute (wpm). The rates for TED talks and radio monologues are at first sight noticeably faster than for lectures, but there is considerable variation in the range of speeds used by different speakers.

Speech event	No. of speakers	Mean wpm	Range	Source
TED talks	9	163	133-188	Dlugan (2012)
Scripted radio monologues	19	160.4	134-195	Tauroza & Allison (1990, p.98)
Lectures in the UK *	30	149.7	58-205	Nesi (2001, pp.207-208)
Lectures to non-native speakers**	22	141.7	102-199	Tauroza & Allison (1990, p.98)

Table 1: Comparison of different categories of monologue.

*averages for whole lectures, including periods of silence for reading calculations or writing
**based on samples of continuous delivery

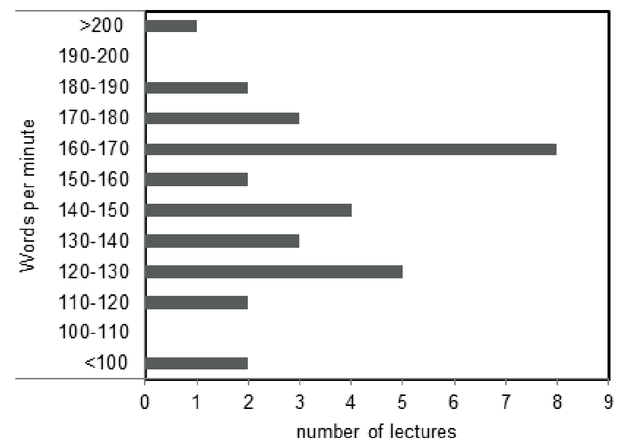


Figure 2: Distribution of BASE corpus lectures by speed (data from Nesi, 2001, pp.207-208)

Speed may be influenced by the purpose of the speaker. A breakdown of Nesi's data from the BASE corpus (2001) shows that lectures mainly range in speed from 110-190 wpm, with a large group falling between 160-170 wpm (Figure 2). However, Nesi notes that faster lectures were sometimes delivered by guest lecturers outside the main academic programme, or when the lecturer seemed not to expect notes to be taken. Nesi also notes examples of sharp changes of pace between sections of lectures containing anecdotes or background information, and sections where note-taking was expected. Although other lecturers in Nesi's sample still expected note-taking whilst speaking at a challenging pace, it would seem that 130-150 wpm could fall within the normal range for lectures where students are expected to identify and record key information.

Once the target speed for the examination lecture has been determined, this can be controlled both during and after the session when the lecture is formally recorded. The speaker can be given guidance about both overall speed and the length of pauses. Pauses should occur naturally at the ends of clauses or sentences, with longer pauses where there are changes of topic. The longer the pauses, the more time listeners have to process ideas. For scripted lectures, speed can also be adjusted by trialling an extract of 100 words or so, and calculating from this whether the speaker needs to go more slowly or quickly for the second take.

Lecture speed can also be adjusted by editing the recording using software such as Audacity®, which is freely downloadable1. The 'Tempo' function of Audacity® can be used to change the pace without distorting the pitch of speakers' voices, and pauses can also be lengthened by pasting in a quarter second of recorded silence. Buck claims tracks can thus be slowed by up to 20% without sounding unnatural (2001, p.186), but these are time-consuming measures which may be avoided by giving tactful feedback to speakers while recording.

The length of the lecture is a further consideration. This will depend on the number of passages included overall in the test and the required number of comprehension items. However, a test of candidates' ability to follow lectures in real life situations is likely to involve a relatively long passage compared with less demanding listening genres; Hughes considers ten minutes or more reasonable (2003, p.164).

TEST TASKS

Having created an appropriate text, test-writers need to decide on techniques and questions to measure the test-takers' ability to understand the lecture. As general rule, however, I would argue that candidates should be required to take notes during a lecture-based test rather than being shown questions before listening. H.D. Brown argues that assessment through note-taking has the advantage of authentically reproducing the experience of classroom lectures (2004, p.136). It is also a thorough measure of comprehension which allows candidates to demonstrate global understanding and ability to identify main arguments and key supporting information. By contrast, pre-questions may have the disadvantage of encouraging test-takers to listen selectively for details, which might be appropriate behaviour for administrative announcements, but not generally for lectures. The authenticity argument also suggests that the recording should be played only once, reproducing the experience of real life lectures. It should be acknowledged, however, that note- taking tasks will be unfamiliar to most students, so adequate classroom practice is essential to enable them to cope with the cognitive demands of simultaneously writing and listening (Carrell, 2007, p.45; Hughes, 2003, p.168).

Candidates' notes can be evaluated in three main ways: by scoring the raw notes, by setting tasks requiring reformulation or transfer of noted information, or through comprehension questions seen after hearing the lecture. There is evidence that raw notes reliably reflect candidates' listening competence. Song (2012) evaluated notes taken by candidates during an 8-minute lecture in comparison to their performance on short answer questions completed after listening. She found the raw notes were good indicators of competence, particularly when assessed in terms of candidates' ability to identify and record main topics (as opposed to minor details), and their ability to organize notes to show hierarchical relationships between ideas. Evaluating raw notes also allows candidates to be given credit for all key information they record, whereas comprehension questions may not cover every noteworthy point. In addition, if training in note-taking techniques has been a teaching focus, this could be factored into scoring. A test task developed by Kahn, for example, includes marks for visual lay-out and use of symbols and abbreviations as well as for accurate recording of main ideas (2002, cited in H.D. Brown, 2004, p.136).

The main drawbacks of assessing notes directly are increased marking time and reduced reliability in scoring (H.D. Brown, 2004, p.136ghes, 2003, p.168). Test-writers must prepare a full list of points to be identified in candidates' notes and the marks to be awarded, possibly including information meriting the award of partial credit. However, since it is virtually impossible to anticipate every acceptable wording of key points, marker subjectivity cannot entirely be avoided. It may not be a coincidence that the three examples of formal assessment of raw notes that I have encountered professionally were all scored by single markers to avoid inconsistency. Another argument against scoring notes directly is that note-taking methods and even the use of L1 are matters of individual freedom, and the comprehensibility of notes should not therefore be subject to external evaluation. However, if note-taking methods have been a focus of teaching, students should be aware of the need to record information clearly, and as Song points out, the seriousness of the test situation should motivate them to take notes relatively assiduously (2012, p74). In addition, students should be allowed to check and edit their notes before final submission, which would require no more time than completing alternative assessment tasks.

The second option, evaluating notes through a reformulation task, has the potential advantages of guiding students to present noted information in a clear format for a marker whilst also retaining the chance to gain credit for all information they have understood. Reformulation can take the form of an extensive writing task, such as a summary which can be scored according to inclusion of main points and/or supporting examples specified in the answer key. A more complex alternative is an integrated writing task requiring candidates to draw on both the lecture and some other input, typically a reading text. Examples include the University of Illinois English Placement Test (Cho, 2003) and the first writing task in the TOEFL iBT (ETS, 2015, pp.24-25). Extensive writing tasks are typically used to assess writing competence as well as ability to identify and use key information from the input. By contrast, candidates could also use their notes to complete an outline or a table, which could test identification of main ideas, but require only key words and phrases rather than a coherently written text. As with raw notes, reformulation tasks are relatively time-consuming and subjective to grade, and require preparation of a list of anticipated answers. This will be facilitated by piloting the test thoroughly with colleagues.

Thirdly, notes may be evaluated indirectly through comprehension questions which candidates see only after hearing the lecture and answer with reference to their notes. As Hughes argues, items can be scored more easily and reliably than raw notes (2003: p.168). However, when creating items, question focus must be considered with care, and Hughes emphasizes that questions must be "perfectly straightforward to anyone who has taken appropriate notes" (2003: p168). According to Weir (1993, cited by Rost, 2005, p.174), this is best achieved by basing the questions on a proficient set of notes taken while listening to the lecture, rather than on the full script. To do this, a rough recording can be made for a colleague to take notes from, with a second listening possibly permitted if the note-taker reports any losses of concentration which might occur without the 'authentic' pressure of being in a real examination. This would also allow incomplete phrases or abbreviations in the initial notes to be written more clearly for the test writer to refer to. Alternatively, the outline notes for a semi-scripted lecture could be used as a basis for question content, with revisions to items made after the actual recording if necessary.

The final decision concerns question style. Multiple Choice Questions (MCQs) are a popular method of assessing comprehension, and are easy to mark. However, they are beset by potential errors and traps for test-writers which result in unfair or unreliable items. In addition, they may inflate the grades of students who have better guessing techniques, or ''test-wiseness" (Allan, 1992). Gap-filling items, in which sentences must be completed by one or two words, are easier to prepare. However, similarly to MCQs, there is a risk that answers may be deduced from contextual clues in the sentence, or simply from common sense. To help avoid this, a colleague can be asked to attempt the questions without knowledge of the text, to ensure that answers really are dependent on comprehension of the input. For a thorough discussion of pitfalls to avoid when preparing MCQs and gap-filling questions, including some helpful checklists for item-writers, see J.D Brown (2006).

CONCLUSION

Short lectures are an appropriate text type for academic listening assessments, but need to be prepared in such a way that they contain features of natural speech typical of this genre. Scripted, semi-scripted and genuine live lectures each have potential to achieve this, but test-preparers should take measures to counter the drawbacks of each of these approaches. Note-taking is recommended as the most authentic assessment task, provided candidates have received adequate training in this skill. Test-writers have three main options for evaluating candidates' performance, including scoring notes, different forms of transformation task, and post-listening comprehension items, but they should be aware of the positive and negative implications of each of these methods when deciding which to use in the test design.

REFERENCES

Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. Language Testing, 9(2), 101-119, doi:10.1177/026553229200900201

Brown, H.D. (2004). Language Assessment: Principles and classroom practices. White Plains, NY: Longman

Brown, J.D. (2006). Chapter 3: Developing good quality language test items. In J.D. Brown, Testing in Language Programs: A Comprehensive guide to English Language Assessment (pp. 41-65). Beijing, China: McGraw Hill / Higher Education Press

Buck, G. (2001). Assessing Listening. Cambridge, UK: Cambridge University Press.

Carr, N.T. (2011). Designing and analyzing language tests. Oxford, UK: Oxford University Press.

Carrell, P.L. (2007). Notetaking Strategies and their relationship to performance on Listening Comprehension and Communicative Assessment Tasks. ETS. Retrieved 7 April 2015 from http://www.ets.org/research/policy_research_reports/publications/report/2007/hsle

Cho, Y. (2003). Assessing writing: Are we bound by only one method? Assessing Writing, 8 (3), 165– 191, doi:10.1016/S1075-2935(03)00018-7

Clark, M. (2014). The use of semi-scripted speech in a listening placement test for university students. Papers in Language Testing and Assessment, 3(2), 1-26. Retrieved 8 April, 2015, from http://www.altaanz.org/uploads/5/9/0/8/5908292/clark.pdf

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment: Structured overview of all CEFR scales. Retrieved 7 April, 2015 from: http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Key_reference/Overview_CEFRscales_EN.pdf

Dlugan, A. (2012). What is the average speaking rate? Retrieved 7 April, 2015 from http://sixminutes.dlugan.com/speaking-rate/

ETS (2015). TOEFL iBT® Test Questions (PDF). Retrieved 20 June, 2015 from http://www.ets.org/toefl/ibt/prepare/sample_questions

Flowerdew, J. & Miller, L. (1992). Student perceptions, problems and strategies in second language lecture comprehension. RELC Journal, 23(2), 60-80, doi:10.1177/003368829202300205

Flowerdew, J. & Miller, L. (1996). Lectures in a second language: Notes towards a cultural grammar. English for Specific Purposes, 15(2), 121-140, doi:10.1016/0889-4906(96)00001-4

Flowerdew, J & Miller, L. (1997). The teaching of academic listening comprehension and the question of authenticity. English for Specific Purposes 16(1), 27-46 doi:10.1016/S0889-4906(96)00030-0

Hughes, A. (2003). Testing for language teachers. (2nd Edition) Cambridge, UK: Cambridge University Press

Lee, J.J. (2009). Size matters: an exploratory comparison of small- and large-class university lecture introductions. English for Specific Purposes 28(1), 42-57 doi:10.1016/j.esp.2008.11.001

Lynch, T. (2011). Academic listening in the 21st century: Revising a decade of research. Journal of English for Academic Purposes, 10(2), 79-88, doi:10.1016/j.jeap.2011.03.001

Nesi, H. (2001). A corpus-based analysis of academic lectures across disciplines. In J. Cotterill, J. & A. Ife (Eds.), Language across boundaries (pp. 201-218). London, UK: Continuum.

Rost, M. (2005). Teaching and researching listening. Beijing, China: Foreign Language Teaching and Research Press.

Song, M-Y. (2012). Note-taking quality and performance on an L2 academic listening test. Language Testing, 29(1), 67-89, doi:10.1177/0265532211415379

Tauroza, S. & Allison, D. (1990). Speech rates in British English. Applied Linguistics, 11(1), 90-105. doi:10.1093/applin/11.1.90

Thompson, S.E. (2003). Text-structuring metadiscourse, intonation and the signalling of organization in academic lectures. Journal of English for Academic Purposes 2(1), 5-20, doi:10.1016/S1475-1585(02)00036-X

[1]Audacity® can be obtained from http://audacity.cn.uptodown.com/ (Chinese) or http://audacity.sourceforge.net/download/ (English)

Author Biography

Nigel Dixon is an EAP Tutor and former Examinations Officer at XJTLU. He has worked on test preparation and taught English in China, Indonesia, Turkey, Thailand and the UK, and has recently gained experience as a language test candidate in Chinese HSK exams.

Email:
nigel.dixon@xjtlu.edu.cn

SPEAKING WITH... DR. STUART PERRIN AND MARKUS DAVIS

The Co-Editors of ETiC recently met up with Dr. Stuart Perrin, Director of the Language Centre, and Markus Davis, Short Courses Manager, at Xi'an Jiaotong-Liverpool University (XJTLU) to discuss the National Certificate in Teaching English for Academic Purposes (NCTEAP). After being corrected on their mispronunciation of its acronym (N-C-TEAP, three syllables, the last rhyming with 'deep'), the conversation focused on the background to this newly-created award.

ETIC:
Very briefly, what is the NCTEAP?

MARKUS:
It's a three-week training programme with practical and theoretical elements which is designed to help teachers who are less familiar with delivering English for Academic Purposes (EAP) in a higher education context.

ETIC:
Why have you decided to run this programme now?

STUART:
The NCTEAP is something which has come about because of a change in the way that Chinese universities are looking at their English language provision. Traditionally, the College English Test (CET) constitutes the English language element of a Chinese degree programme. To graduate, a student needs to pass CET 4 or, in many institutions, CET 6. However, there are radical changes unfolding in Chinese higher education at the minute and the role of English language is one of them.

ETIC:
What are the changes exactly?

STUART:
Well, one of the discussions taking place revolves around whether English should be in the curriculum at all. There are those who think there is too much English in the curriculum and that this is having adverse effects on the 'Chineseness' of students. There is some evidence of this influence at high school level where the importance of English in the GaoKao is being downgraded slightly.

Then, there are debates about if English is in the curriculum, what it should look like. At a time when many British universities are moving away from EAP towards things such as Writing in the Disciplines, Writing Across the Curriculum, and Academic Literacies, China is discovering English for Academic Purposes. So, there is a movement within China looking at how that can be incorporated into the curriculum here. Even among the advocates for EAP there are a number of rival groups vying to be at the forefront of the movement. The China English for Academic Purposes Association (CEAPA), which XJTLU is aligned to, is one.

ETIC:
You've mentioned the College English Test (CET) and English for Academic Purposes (EAP) as different routes for the future of English language provision in China. How would they be different?

STUART:
Within a CET context, teachers are essentially preparing their students to pass an exam. So, naturally the focus is very much on the skills needed for that. EAP takes a much more holistic approach to teaching. It is not just about teaching EAP for its own sake, although that is a worthy thing to do, but also about enabling students to apply what they have learnt to their wider studies conducted in English.

ETIC:
You seem to be hinting at the possibility of students having to use English more meaningfully on degree programmes.

STUART:
That is a possibility. I'm currently involved in a project with the Department of Education in Jiangsu Province. They have given me and some other people money to look at and research the concept of an English-led curriculum within the province. Jiangsu scores very highly when it comes to education in China. It's usually ranked first or second in many areas. In terms of internationalisation, however, it isn't ranked quite so high. So, ways of improving its standing in terms of internationalisation are being explored. Partly, that means getting more international students into Jiangsu universities. But another aspect is developing more modules and courses which are delivered in English. What all of this means is that Chinese academics who might end up delivering this new type of curriculum, and especially Chinese teachers of English, will be asked to do things they are not necessarily trained in doing. The NCTEAP is a way of providing this training.

ETIC:
We've talked a lot so far about the Chinese context. Is the NCTEAP only for teachers working in China?

MARKUS:
Not at all. Even though we work in a Chinese context, the lead trainer, Dr. Tim Marr, doesn't. There is a universal element to the training. The delivery of EAP is always context driven. That can be very local, it can be national, or it can be international. Whatever training anyone receives, they have to take it back and contextualise it in their own teaching scenario.

We have spoken to teachers who do work in other places and there is a relevance for them. For example, three teachers from the Philippines came to our National EAP Training Day in November. Stuart mentioned how people who are currently teaching the CET might find that their work is changing, but there are other people overseas who are perhaps teaching things like FCE (one of the Cambridge English exams) who might be looking to get into teaching EAP.

STUART:
To step back a little, another reason we thought about running the NCTEAP was because of the success of the National EAP Training Day. About two hundred people came and although there was an appreciation of what we were doing in terms of how we were training and how we were engaging with the core audience, there also seemed to be a demand for something beyond a Saturday afternoon in Suzhou.

When we did our due diligence, we had a look at what other types of programs such as this were on offer. And actually there isn't that much on offer at all. If you think about the experience of many teachers, even of those working here within the Language Centre, how did people come to work in English for Academic Purposes? Many of them had never been trained specifically in teaching EAP. They were working in a private language school or maybe in a high school and then they found themselves working in a university and they were expected to adapt. So it seemed to us that there was a clear demand not just within China, but generally there is a demand for training in teaching EAP.

ETIC:
With that in mind, could you give us a typical profile of an NCTEAP trainee?

MARKUS:
Well, there potentially could be quite an incredible range, so I wouldn't want to give the impression that there is only one type of career profile which is suitable for this course. To pluck an example out of thin air, you might be looking at someone who has just graduated with an MA TESOL and who has experience of teaching within the private language school sector. In the future, they would like to apply to become a teacher here at XJTLU, but they don't have grounded work experience in EAP. The NCTEAP would provide them with training that other candidates for the same kinds of jobs might not have.

STUART:
Yes, that's quite a good profile. When we put an advert out we get a range of people who are interested in working here. For a Chinese first language speaker this is a particularly attractive job but often they don't have the experience that we would require and a course like this would help to give those people a little bit more experience through the training and subsequently make them a little bit more employable.

MARKUS:
Not only that, but recently I have been talking to administrators of courses on which the students are doing the first part of a British degree in a campus in China which is not owned by the British university. The requirement for tutors working in such an environment is not only to develop language skills but also to develop EAP skills, which we've been talking about. The organisers of those courses are looking around and asking themselves where they can find people who can deliver these courses.

ETIC:
What about a typical profile of a trainer?

MARKUS:
As I mentioned earlier, the lead trainer is Dr. Tim Marr. He has recently co-authored Why Do Linguistics?, an introductory guide to the study of language. More importantly from our point of view, he has extensive experience and expertise in teacher training going back to the 1990s with the British Council in Thailand. More recently he has been director of the MA TESOL programme at London Metropolitan University, working with trainee teachers from all over the world, and in the last few years predominantly from China. His role will be to forge the direction and the philosophy of the training team. Not only will he be a deliverer of training, he will also be the trainer of the trainers, and will be designing the syllabus in conjunction with senior staff here.
We're also looking to utilise the experience and skills of our colleagues who've got feet on the ground. Tim aside, all the other trainers will be employees of XJTLU. They will all be experienced EAP teachers who have worked at XJTLU for at least one academic year. I expect there to be a wide range in terms of ethnic background, gender and age. We have a very diverse group of teachers here and the training team will reflect that.

ETIC:
Well, thank you both for your time and for sharing your thoughts. We wish you all the best with the NCTEAP.

The NCTEAP runs between 25th July and 12th August 2016. For more information and to apply, contact ncteap@xjtlu.edu.cn

XJTLU NCTEAP

National
Certificate in
Teaching
English for
Academic
Purposes

A TEACHER TRAINING COURSE FOR ENGLISH FOR ACADEMIC PURPOSES (EAP) PROFESSIONALS. CREATED BY XJTLU, THE COURSE ADVANCES THE VERY BEST PRACTICE IN EAP IN HIGHER EDUCATION IN CHINA AND BEYOND. RECOGNISED BY EMPLOYERS, IT IS A PRACTICAL CERTIFICATE THAT WILL HAVE A DIRECT IMPACT ON A TEACHER'S UNDERSTANDING OF EAP AND THEIR CLASSROOM PRACTICE.

www.xjtlu.edu.cn
ncteap@xjtlu.edu.cn

Xi'an Jiaotong-Liverpool University
西交利物浦大学

THE NCTEAP COURSE

On the course you will receive practical training in how to design and deliver an EAP curriculum at a Higher Education institution. You will observe our presessional classes and reflect upon your own practice. You will also be given a brief theoretical grounding in the latest sociolinguistic ideas about the role of English in China and around the world. Some examples of the certificate's content are:

- Teaching the four skills (reading, writing, listening and speaking) in an EAP context;
- Syllabus design: setting learning outcomes;
- Designing EAP assessments;
- Preparing students for lectures and seminars;
- The student-centred classroom;
- Working with academic departments;
- New technologies: using audio-visual equipment and VLEs to support learning;
- Giving feedback to students, including effective error correction.

The training sessions will take place in the morning and early afternoon, leaving time for a range of social activities to enjoy in Suzhou and its surrounding areas.

COURSE DATES

The course is an intensive three weeks, with 20 hours a week of training sessions: Monday 25th July to Friday 15th August 2016

COURSE FEES

The fees listed here are for the NCEAP course tuition, and upon successful completion, the NCTEAP certificate. Accommodation and living expenses are not included.
Full Tuition Fees: 10,000 RMB

EARLY REGISTRATION DISCOUNT

If you apply and pay before 1st June 2016 then you will receive a 20% discount.
Early Registration Discount: 8,000 RMB

INSTITUTIONAL SPONSORSHIP

If an institution, such as a university or college, is sponsoring teachers to take the NCTEAP then for every three teachers that take the certificate we will offer one free place.
Four Sponsored Teachers: 30,000 RMB

The Early Registration Discount also applies to institutional places. If four places from the same institution are applied and paid for by 1st June 2016 then the total discount will be 40%!
Early Registration Discount (Four Sponsored Teachers): 24,000 RMB

ENGLISH LANGUAGE TEACHER MOTIVATION AND TURNOVER IN A PRIVATE LANGUAGE INSTITUTION IN CHINA

Ross Thorburn

ABSTRACT
The purpose of this study was to investigate the influences and relative importance of different motivation and hygiene factors with regards to teachers' decisions to either renew their teaching contracts or resign at a private language institution in China. Data was collected from a survey of 278 teachers between October 2014 and April 2015. Findings indicate that teachers were primarily motivated to renew their contracts by career opportunities, enjoyment of the work itself and growth. Teachers were primarily motivated to leave their positions due to dissatisfaction with company policies and poor relationships with their managers. Teachers commonly cited "China" as a reason for both renewing their contracts and for resigning.

摘要
本项研究的目的是调查当在中国的教师需要做出续约或辞职的决定时，哪些是相关和重要的影响因素，以及积极推动和消极维持的因素。从2014年10月至2015年4月，在针对278教师的问卷调查数据显示，最主要影响教师选择续约的原因有职业发展，对工作本身的喜爱以及在工作当中的成长；同样，最主要影响教师选择离开的原因有对公司制度的不满意，以及和直属主管之间的不良关系。大部分教师也会引用“中国”作为续约或辞职的同等因素。

INTRODUCTION
This article will review issues related to teacher turnover with a specific focus on China, review relevant literature on the subject, consider data from several hundred teachers at a private language institution in China, attempt to explain the results and consider the implications for other language teaching institutions in China.
Teacher turnover is a challenge across the world. In mainstream education, approximately half of all teachers leave the teaching profession within five years of joining (Ingersoll & Smith, 2003). This creates a problem. Even during a time when there is a perceived move towards a reliance on technology in education, teacher effectiveness remains the most important in-school factor affecting student learning (Rivkin, Hanushek & Kain, 2005). Teachers who remain in the profession improve in effectiveness in their first few years (Henry, Bastian & Fortner, 2011). In short, teacher turnover harms student learning.

In China the problem is arguably even greater. Teacher turnover in China is so high that approximately 100,000 foreign teachers and experts are recruited each year to work in mainland China (Swanson, 2013). If we accept that effective teachers are a key factor in student learning; that the longer teachers work as teachers the more effective they become; and that there are around 300 million English learners in China (Swanson, 2013) then the issue of teacher turnover in China is one of the more important contemporary issues in the language teaching industry and one which deserves our attention.

MOTIVATION
Motivation plays a key role in employee turnover. The more satisfied employees are in their jobs, the less likely they are to leave (Heskett, Jones, Loveman, Sasser & Schlesinger, 1994). The corollary to this is that employees who are demotivated, for example as a result of perceived inequities in the workplace, will take action to respond to their demotivation, including the possibility of resignation (Robbins, 2011).
In Herzberg's (1987) influential work on motivation, he hypothesized that there were two forms of motivation. These were termed "motivators" and "hygiene factors". Herzberg found the presence of motivators generally resulted in job satisfaction. Motivators included achievement, recognition, the work itself, responsibility, career advancement and growth. The absence of hygiene factors generally caused dissatisfaction at work. Hygiene factors included company policies, supervision, relationships with supervisors and peers, work conditions and salary.
More recently, Groysberg, Nohria and Lee (2008) have hypothesized that humans are motivated by four main drives:

- The drive to acquire (salary, status)
- The drive to bond (relationships with peers and the organization)
- The drive to comprehend (contributing, doing meaningful, interesting and challenging work)
- The drive to defend (security, "organizations…which have clear goals and intentions, and that allow people to express their ideas and opinions" (p. 3).

Much has been written about the importance of salary in job satisfaction and employee motivation. Judge, Piccolo, Podsakoff, Shaw and Rich (2010) argue that while pay is motivating for many individuals, high salaries do not result in a satisfied workforce. Deci and Ryan (1985) suggest that extrinsic rewards cause demotivation and dissatisfaction to individuals. Employees are more likely to enjoy their jobs if they focus on the work itself, and less likely to enjoy their jobs if they are focused on money (Chamorro-Premuzic, 2013). Furthermore, offering bonuses or pay increases is an inefficient method of decreasing employee turnover (Allen, 2008).
Motivations are also believed to vary from one generation to another. Generation Y employees (born between 1980 and 1994) are thought to be more motivated by advancement and free time and less motivated by responsibility and compensation compared with their Generation X counterparts (born between 1965 and 1979) (Barford & Hester, 2011).
Hockley (2006) conducted research with language teachers with the aim of identifying which motivation factors were most important to teachers of English as a second or foreign language. 105 teachers located in 12 different countries working in 91 different schools (many of which were based in the Middle East) were surveyed using factors derived from a survey by Montana and Charnov (2000, as cited in Hockley, 2006). Hockley found that the following factors were the most important in motivating language teachers:

- Respect for me as a person
- Good pay
- Getting along well with others on the job
- Opportunity to do interesting work
- Feeling my job is important
- Opportunity for self-development and improvement (p. 11).

AIM
This research set out to discover the main reasons which caused teachers to renew their annual contract or resign and compare these with motivational factors found by Herzberg (1987), Hockley (2006) and Groysberg et al. (2008). Reasons for similarities and differences between their results and those obtained in this research will be subsequently discussed as will the implications for language schools in China.

METHOD
SAMPLE AND PROCEDURES
Between October 2014 and April 2015, 278 completed responses were received from teachers and Directors of Studies in 16 cities in China working for one private language teaching institution with over one hundred schools throughout China. Teachers in this institution had an average

age of 29, with 83% of teachers born post 1980 (Generation Y) and 17% born between 1965 and 1979 (Generation X). 50% of the respondents taught young learners, 50% taught adults, 93% were expatriate teachers (most commonly from the USA, UK, Canada and Australia) and 7% were Chinese teachers of English. All teachers were offered a bonus and/or a salary increase to encourage them to renew their contracts. Completed surveys were received from teachers who had:

- renewed their contract (188 responses – 60% response rate)
- resigned (90 responses – 30% response rate).

Surveys were emailed to teachers within one week of their announcing their decision to renew their contract or resign. Teachers selected the primary reason for making their decision to either renew their contract or resign from a multiple choice list of 13 options. Open-ended responses were invited using a space for additional comments. Teachers were asked if they would recommend their friends to work at this school on a scale of 1-5 and how long they initially intended to work at the school. Teachers who resigned were also asked about their future career plans.

DATA ANALYSIS

For ease of analysis and comparison, similar factors were used in this survey as were used by Herzberg (1987), as Herzberg's theory is still considered to be valid more than 50 years after it was first proposed (Jones & Lloyd, 2005). Three options were added to the survey. These were "the schedule" (combined with "company policies" in the results section), "training & development," (combined with "growth" in the results section) and "the desire to stay in/ leave China" (as Herzberg's original research was carried out with domestic, not expatriate workers).

From the results of these surveys, it is possible to determine the most important factors for EFL teachers in China in relation to renewing a contract and resigning. It is also possible to make recommendations to language schools in China on factors which will help to reduce teacher turnover.

RESULTS

MOTIVATORS

The most common reasons teachers chose as reasons for renewing their contracts were:

- Career opportunities
- Growth, training and development
- The work itself
- Wanting to stay in China.

HYGIENE FACTORS

The most common reasons teachers chose as the primary reason they resigned were:

- Company policies (including "schedule")
- Relationships with managers
- Wanting to leave China.

These are illustrated in Figure 1.

DISCUSSION & IMPLICATIONS

The most commonly chosen motivation and hygiene factors will be discussed.

CHINA

Perhaps unsurprisingly, China was the only factor significant in teachers' decisions for both renewing their contracts and resigning. Indeed, "the desire to travel and live and work in different countries and cultures is often a strong motivating force [for teachers]" (Hockley, 2006, p. 5). However, the uncertainty and ambiguity of living and working abroad can also cause stress (Mezias & Scandura, 2005), which among other factors leads to between 20% and 40% of expatriate workers returning home before completing their expected tenure abroad (Black & Mendenhall, 1989; Kim & Slocum, 2008; Mendenhall, Dunbar & Oddou, 1987).

In this study, several teachers commented negatively about air pollution in China, saying "Beijing has been difficult to live in. I am constantly coughing and often feel sick." "I suffer from asthma and Beijing's air quality has made me very sick over the past couple of months." Other teachers appeared to have difficulty adapting to living abroad, commenting, "China is very difficult to navigate as a foreigner" and "I felt completely helpless in China. I wish there had been someone who was there to help us along the way to offer advice." Some teachers who cited China as a reason for resigning were doing so as part of a plan to return home, for example "I'm leaving because China cannot offer me what I need in this stage of my life."

Many of the comments related to China were positive. Teachers who renewed their contracts commented "Living in China is a fun experience" and "I enjoy the opportunity to enhance my knowledge of China and its people."

These results suggest that adapting to living in a new country is an important factor in foreign teacher turnover. Helping employees engage in the host culture, providing training on overcoming culture shock, allowing extended time off to facilitate travel home, providing language lessons and assistance in apartment hunting or providing housing to teachers are all steps which schools may take to help teachers overcome the challenges of moving to a new country.

GROWTH, TRAINING & DEVELOPMENT

Over 40% of the respondents who renewed their contracts were found to have undertaken a training course within the past year compared with just 12% of teachers who resigned. In recent research conducted by Universum, a Swedish consultancy which specializes in talent acquisition and brand building, 45% of Generation Y respondents emphasized the importance of learning and

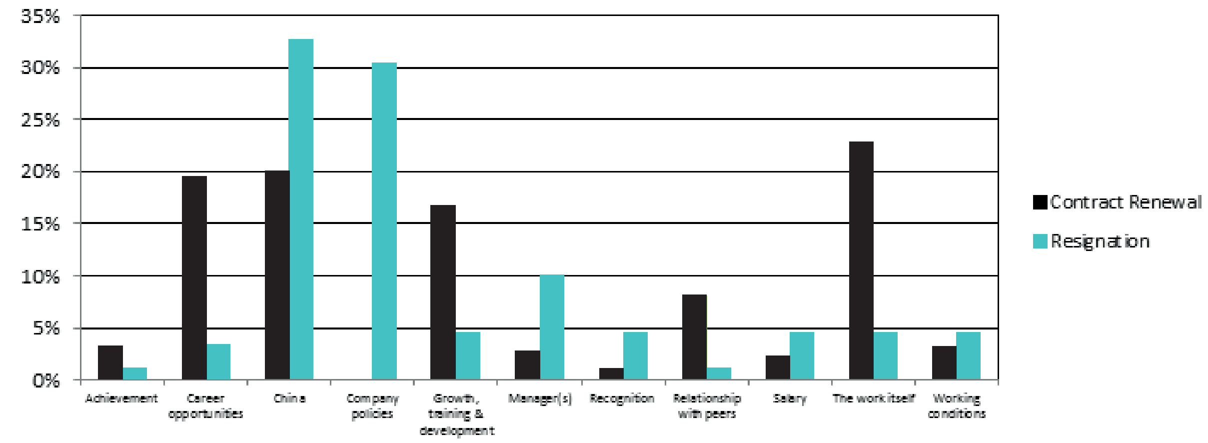


Figure 1: Primary reasons teachers cited for contract renewal / resigning

developing new skills (Dill, 2014). The results here reflect the importance of teacher development in retaining teachers.

Teachers who chose "growth, training and development" gave high Employee Net Promoter Scores (eNPS), a popular method used by employers to evaluate employee satisfaction (Reichheld, 2003) compared with other teachers. This was true for both expatriate and Chinese teachers. Thus, teachers who are primarily motivated by their own professional growth are more likely to be loyal employees, stay longer and promote their place of work to their friends (Kaufman, Markey, Burton & Azzarello, 2013).

This option was chosen more frequently by Chinese teachers than by their expatriate counterparts as a reason for choosing to renew their contracts. This factor may have been more commonly selected by Chinese staff as they intended to stay in the teaching profession for longer than their expatriate counterparts. 67% of the Chinese teachers who renewed their contracts said they initially planned to work at this private language institution for more than one year when they started, compared with just 35% of expatriate teachers. It also could be argued that, as none of the Chinese teachers surveyed were motivated by working in China, other motivators took on greater relative importance.

CAREER OPPORTUNITIES

Previous research by Hockley (2006) suggested "career opportunities" were not important to language teachers, going so far as to say "most teachers I have talked to have no great interest in becoming managers" (p. 3). The results here are quite different: career opportunities being the primary reason one in five teachers renewed their contracts. This may be a result of Generation Y employees placing more importance on personal growth than the Generation X employees surveyed by Hockley a decade ago. Recent research conducted by Universum showed that 40% of Generation Y respondents "list their biggest fear as becoming trapped in a job with no chance for development" (Dill, 2014, p. 2). "Forty-one percent of respondents said taking on a leadership or management role was 'very important to them'" (Dill, 2014, p. 2).

Career opportunities are common in some privately owned language schools in China, where teachers can find themselves in management positions within a relatively short space of time, often as a result of teacher turnover. This contrasts with language schools in other countries, where "in most cases there exists neither the opportunity nor the desire for promotion in the traditional hierarchical sense of the word" (Hockley, 2006, p. 3).

Groysberg et al. (2008) included improving one's social status as part of the drive to acquire and Herzberg (1987) found "advancement" to be the fifth most important motivation factor which is consistent with the findings in this study.
Company policies
Herzberg found "company policy and administration" to be the main source of "extreme dissatisfaction" in his research. For English teachers in China, the company policies which caused dissatisfaction were generally related to scheduling. Teachers who selected this as their reason for leaving said "I didn't want to work weekends anymore," "[I have] unsocial working hours" and "I hope we could take unpaid leave for attending training sessions instead of using our annual leave." This issue may be particularly prevalent in private education companies in China where the majority of classes

take place during evenings and weekends. These teachers' attitudes reflect the importance of free-time for Generation Y employees (Barford & Hester, 2011).

The other company policies referred to were related to the for-profit nature of the business, with teachers commenting "policies are driven by the need for more profits rather than what is good for the employees or students" and "there is too much focus on money and not enough on academic integrity." Realistic job previews can decrease turnover of employees by up to 50% (Suszko & Breagh, 1986). In this case, setting clearer expectations during the recruitment process regarding the profit-driven nature of privately owned language schools in China and emphasizing the unsociable working hours of teachers in private language institutions may have helped to avoid dissatisfaction with these policies.

THE WORK ITSELF

Hockley (2006), Herzberg (1987) and Groysberg et al. (2008) all previously identified "the work itself" (or "the opportunity to do interesting work" or jobs which are "meaningful, interesting and challenging") as a motivator. Here, this was the most common reason teachers cited for renewing their contracts. Teachers of adults chose this option three times as often as teachers of young learners did. It is possible that working with young learners and coping with behavior problems is often overwhelming for new teachers and prevented them from choosing this reason.

Chinese teachers were twice as likely to choose "the work itself" as a reason for renewing their contracts as expatriates were. Twice as many Chinese teachers who renewed their contracts initially planned to stay with the company for more than one year compared with the expatriate teachers surveyed. This suggests that these teachers were more confident in teaching as a longer-term career option compared with the expatriates surveyed. It is therefore not surprising that these teachers were also motivated to renew their contracts by their enjoyment of teaching. However, that Chinese teachers cited this factor more than expatriate teachers may also indicate that other options which were important to expatriate teachers, such as career opportunities, may be perceived to be less easily obtainable for Chinese staff. While 22% of the expatriate teachers cited "career opportunities" as their primary reason for renewing their contracts, the same was true for merely 9% of Chinese teachers.

RELATIONSHIPS WITH MANAGERS

Herzberg (1987) also found relationships with managers (or "supervisors") and "supervision" to be two of the main hygiene factors identified in his research. Groysberg et al. (2008) acknowledged that direct managers were as important as organizational policies in employee motivation.

Teachers who selected their manager as the main reason they left the company were less likely to recommend the school to a friend compared with teachers who cited wanting to leave China as their primary reason for resignation. Negative relationships between managers and teachers can therefore potentially harm recruitment initiatives by speaking negatively about the company in future. The majority of the teachers who cited their managers as the reason they left indicated that they intended to remain as teachers in China.

In the multinational and multicultural work context, trained, competent managers are essential. Training for managers on employee motivation, intercultural communication as well as stressing the importance of open dialogue with employees can all help to decrease teacher turnover.

SALARY

Salary did not appear to be a major contributing factor in teachers' decision making and motivation in this study. This contrasts with other findings in the literature. Hockley (2006) found salary was the second most important factor in teacher motivation. Groysberg et al. (2008) also recommended satisfying the drive to acquire through a reward system, which includes salary.

However, the results do agree with Judge et al. (2011) who recently reviewed 120 years of research on the subject and found minimal correlation between pay and job satisfaction noting, "relatively well-paid samples of individuals are only trivially more satisfied than relatively poorly paid samples" (p. 162). Other research has shown that intrinsic job characteristics are a better predictor of job satisfaction than salary (Judge & Church, 2000).

It is possible that teachers who are primarily motivated by salary choose to work for companies or in countries with higher starting salaries, for example in the Middle East. Demographics may again partially explain these differences; the majority of participants in Hockley's (2006) study were more experienced, more highly paid and older than the teachers surveyed here. Generation Y employees are believed to be less motivated by compensation than their Generation X counterparts.

CONCLUSIONS

For the teachers surveyed, career opportunities and growth, training & development were the main drivers for over one third of the teachers to renew their contracts. Investing in training and development to help teachers grow professionally is clearly vital for Generation Y English language teachers and a more effective retention tool than salary increases or renewal bonuses. Language schools in China should ensure organizational rewards based on performance are factors which motivate teachers, i.e. access to and sponsorship for training and career advancement and not simply performance related pay.

Little seems to have changed in terms of what demotivates since Herzberg's (1987) research was conducted. Supervision and company policy both featured prominently in Herzberg's research and these factors also feature prominently here. There are implications for schools in terms of job design (sociable working hours, unpaid leave to allow longer vacations and visits home) and the selection and training of interculturally competent managers, the absence of which not only results in teacher turnover, but also in ex-employees who will be likely to speak negatively about their former school and thus harm future teacher recruitment. Setting clear expectations as part of the recruitment process about unsociable working hours and other company policies could also help avoid unnecessary dissatisfaction and decrease turnover.

China was the single most important factor in teachers' decision to either renew their contract or resign. Meeting teachers' most basic needs when moving to a new country, providing assistance to them in finding an apartment, overcoming culture shock, learning the local language and assisting them in building a social network when they arrive could all contribute to decreasing teacher turnover. The English language teaching industry in China is worth $2.1 billion a year (Swanson, 2013). Investing more in teachers' growth, development and general wellbeing will pay dividends not only for individual schools but for the 300 million English learners in China.

REFERENCES

Allen, D. G. (2008). Retaining talent: A guide to analyzing and managing employee turnover. Retrieved from http://www.shrm.org/about/foundation/research/documents/retaining%20talent-%20final.pdf

Barford, I., & Hester, P. (2011). Analyzing Generation Y workforce motivation. Defense AT&L, 40(2), 36-40.

Black, J. S., & Mendenhall, M. (1989). A practical but theory-based framework for selecting cross-cultural training methods. Human Resource Management, 28(4), 511–539.

Chamorro-Premuzic, T. (2013, April). Does money really affect motivation? A review of the research. Retrieved from https://hbr.org/2013/04/does-money-really-affect-motiv.

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. New York: Plenum.

Dill, K. (2014). 7 things you (probably) don't realize about your millennial employees. Retrieved from http://www.forbes.com/sites/kathryndill/2014/11/13/7-things-you-probably-dont-realize-about-your-millennial-employees/

Groysberg, B., Nohria, N., & Lee, L. (2008). Employee Motivation: A Powerful New Model. Harvard Business Review, 86(7/8), 78-84.

Henry, G. T., Bastian, K. C., & Fortner, C. K. (2011). Stayers and leavers: Early-career teacher effectiveness and attrition. Educational Researcher, 40(6), 271–280.

Herzberg, F. (1987). One More Time: How Do You Motivate Employees? Harvard Business Review, 67(5), 109-120.

Heskett, L., Jones, T., Loveman, G., Sasser, E., & Schlesinger, L. (1994). Putting the service-profit chain to work. Harvard Business Review, 72(2), 164-170.

Hockley, A. (2006, March). What makes teachers tick? IATEFL Leadership & Management SIG – Newsletter, 37, 10-13.

Ingersoll, R. M., & Smith, T. M. (2003). The wrong solution to the teacher shortage. Educational Leadership, 60(8), 30–33.

Jones, N. B., & Lloyd, G. C. (2005). Does Herzberg's motivation theory have staying power? Journal of Management Development, 24(10), 929-943.

Judge, T. A., & Church, A. H. (2000). Job satisfaction: Research and practice. In C. L. Cooper & E. A. Locke (Eds.), Industrial and organizational psychology: Linking theory with practice (pp. 166–198). Oxford, UK: Blackwell.

Judge, T. A., Piccolo, R., Podsakoff, N., Shaw, J., & Rich, B. (2010). The relationship between pay and job satisfaction: A meta-analysis of the literature. Journal of Vocational Behavior, 77(2), 157-167.

Kaufman, J., Markey, R., Burton, S., & Azzarello, D. (2013, December). Who's responsible for employee engagement? Retrieved from http://www.bain.com/publications/articles/whos-responsible-for-employee-engagement.aspx.

Kim, K., & Slocum, J. W. (2008). Individual differences and expatriate assignment effectiveness: The case of U.S based Korean expatriates. Journal of World Business, 43, 109–126.

Mendenhall, M. E., Dunbar, E., & Oddou, G. (1987). Expatriate selection, training and career-pathing: A review and critique. Human Resource Management, 26(3), 331–345.

Mezias, J. M., & Scandura, T. A. (2005). A needs-driven approach to expatriate adjustment and career development: A multiple mentoring perspective. Journal of International Business Studies, 36, 519-538.

Reichheld, F. F. (2003). The One Number You Need to Grow. Harvard Business Review, 81(12), 46-54.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. Econometrica, 73(2), 417–458.

Robbins, S. P., Judge T. A., Millet, B., & Boyle M. (2011). Organizational behaviour. 6th ed. Frenchs Forrest: Pearson Australia.

Suszko, M. J., & Breagh, J. A. (1986). The effects of realistic job previews on applicant self-selection and turnover, satisfaction, and coping ability. Journal of Management, 12, 513-523.

Swanson, T. (2013).). Great Wall of Numbers: Business Opportunities & Challenges in China [Kindle version]. Retrieved from www.amazon.com.

Author Biography

Ross Thorburn has worked in the language teaching industry in China for the past nine years. He currently works as a Training & Development Manager, decreasing staff turnover through providing high-quality training and internationally recognized qualifications to both teachers and managers.

Email:
ross.thorburn@yahoo.com

THE USE OF IB ASSESSMENT RUBRICS; TEACHER AND STUDENT VIEWS COMPARED

Douglas Glenn & Graham Morton

ABSTRACT
This article analyzes the results of two surveys, one given to teachers, the other to students, at an international school in Suzhou, China. The surveys were designed to determine the perceptions of the usefulness and fairness of rubrics for assessment and feedback. The results show that around 50% of both groups believed that rubrics can be difficult to understand and fail to show students how to get a good grade. The article concludes that rubrics, which are commonly used in schools, fail to provide the quality feedback which face-to-face interaction and detailed written feedback offer.

摘要
本文分析了对中国苏州一所国际学校教师和学生分别所做的两项调查的结果。调查设计旨在了解师生如何看待‘评价和反馈量表’的有效性和公平性。调查结果表明，两组对象中各有50%左右的人群认为该量表很难理解，不能向学生展示如何取得好成绩。文章最后指出尽管该量表在学校普遍使用，但与传统名面对面交流和详细的书面反馈相比，不能提供高质量的反馈信息。

INTRODUCTION
Authentic assessment is an alternative to standardized testing. Wiggins (1993) defines authentic assessment as tasks based on "engaging and worthy problems or questions of importance". Authentic assessment is also referred to as "direct assessment" or "performance assessment" because students show what they can do, rather than what they can recall (Mueller, 2014). An authentic assessment is based on a product such as a research essay, physical creation, or performance (such as an oral presentation or a skit) and generally involves the use of a rubric (also referred to as a 'band descriptor' or 'rating scale'), a grid delineating levels of performance across a range of criteria representing the major aspects of the assignment. Brookhart (2013) suggests that clear expectations conveyed through rubrics will not only help students succeed by crystallizing the key concepts of the unit, but can also serve as tools for both formative, peer, and self-assessment. Additionally, rubrics are intended to clarify unit or course objectives in a way that is less time-consuming for both students and teachers.

The authors' school, at which the research for this article was conducted, is a full International Baccalaureate (IB) school located in Suzhou, China. Students attending this school follow the Primary Years Program (PYP) in Grades 1-5 (ages 6-11), the Middle Years Program (MYP) in Grades 6-10 (ages 11-16), and the Diploma Program (DP) in Grades 11 and 12 (ages 16-18). Approximately 80% of the 1200 students enrolled speak English as a second language.

Assessment at the school follows IB guidelines, which for authentic summative assessments require the use of IB-standardized evaluation criteria and subject-specific rubrics. The IBMYP, which is the area of focus in this investigation, allows some latitude in the way each course is taught. Although details of the evaluation criteria and an assessment rubric must be included in all summative assessments, neither is mandated for formative assessments. Graded assessments are returned to the students and include feedback consisting of either a rubric only, a rubric with relevant sections highlighted, written comments, oral feedback, or a combination of these, depending on the task and teacher.

A typical MYP student at the authors' school encounters approximately 120 summative assessment rubrics during one academic year. Students study eight subjects with each subject programme consisting of four to five units of study on different topics delivered over the academic year. These units are assessed summatively using IB rubrics based on IB criteria, typically four for each subject area. The school requirement is to assess each of these criteria twice per semester, or four times per academic year. As a result, these MYP students can be considered to be knowledgeable end-users of rubrics. Similarly, the teachers encounter a large number of rubrics each year and can also be considered 'expert' users of rubrics.

This wealth of experience provided the initial spark for the investigation. The researchers were curious to know what the student and teacher views of IB assessments rubrics are. There were three main questions being considered:

1. Do the teachers and students share similar views about IB rubrics?
2. Are IB rubrics seen to be a fair way of grading assignments?
3. Do they provide valuable feedback?

Through this investigation, the researchers hoped to answer these questions and obtain any additional insights these rubric users could offer. They also believed that these answers and insights could be of interest to other rubric users at all levels of education.

THE USE OF RUBRICS
A rubric is a simple indicator of what performance measures will be considered and how success is defined. Rubrics can be divided into two main varieties: holistic and analytic. A holistic rubric provides feedback on student performance as a whole. Analytic rubrics, on the other hand, consider the various aspects of assessment, such as content, presentation, and use of language, separately (Center for Advanced Research in Language Acquisition, 2015). Because analytic rubrics provide more detail to students, they are more widely used. A third type of rubric, task-specific, is unique to an assessment task, and each different task requires a different rubric.

The standard analytic rubric, which is almost universally used (DePaul University Teaching Commons, 2015), is in grid form. The rows (usually four or five) of the rubric list the criteria to be assessed, while the columns (again usually four or five) describe the levels of success using terms such as "exceeds expectations", "exemplary", "needs work", or "novice". By expressing assignment requirements in terms of the final assessment criteria, rubrics represent a compact means to inform students about what is expected of them. The expectation is that students will frequently refer to it in order to produce higher-quality work.

ADVANTAGES OF RUBRICS FOR TEACHERS
A survey of the educational literature reveals the following advantages of rubrics for teachers:

1. Rubrics help teachers tailor their teaching to student learning goals (Cooper and Gargan, 2009)
2. Rubrics represent consistency, standardizing assessments across different teachers as well as longitudinally across time (Valenza, 2000)

3. Using rubrics means that all students are evaluated the same way (Brookhart, 2008)
4. Rubrics facilitate the task of grading, replacing individual comments with a common framework (Brookhart, 2008).

ADVANTAGES OF RUBRICS FOR STUDENTS

The following advantages of rubrics for students have also been noted:

1. Rubrics are an instant organizer, allowing students to see both unit objectives and teacher feedback side by side (Brookhart, 2008).
2. Students can use rubrics to evaluate themselves at any point during the progress of a unit to see where their strengths and weaknesses lie (Andrade, 2007).

For these reasons, rubrics are highly attractive to many teachers, allowing them to convey their expectations to their students simply and concisely. To these teachers, rubrics embody the learning standards in a fundamental way, making it clear that these standards apply to every student equally. Their compact nature makes them relatively easy to construct as well. Yet, it is as a marking tool that rubrics are viewed to be a real time-saver, with their array of pre-written, standardized comments.

DISADVANTAGES OF RUBRICS

Education researchers have also noted the following disadvantages of rubrics:

1. Too often, rubrics are constructed without consideration of their validity or their reliability (Andrade, 2005).
2. Rubrics lack objectivity because the descriptors used are vague and/or open to interpretation (Popham, 1997).
3. Rubrics also fail to be objective because the scoring and weighting of the criteria being measured are arbitrary (Newkirk, 2000).

Jonsson and Svingby's (2007) review of articles addressing the degree to which rubrics assessed performance found that well-constructed analytic rubrics increased the reliability of assessment, but that rubrics used in isolation for performance assessment were not sufficiently valid tools.

Taking another perspective, Kohn (2011) criticizes rubrics not on the basis of their shortcomings per se, but because they have been so widely adopted. In Kohn's view, when rubrics are given to students at the beginning of a unit to be used as guides, attention instantly shifts towards assessment at the expense of actual learning. In other words, students begin to self-consciously concentrate on what the teacher wants them to produce, rather than what they can independently discover.

METHODOLOGY

Students and teachers were asked to evaluate the effectiveness of the analytic rubrics they used. A questionnaire was chosen as the most practical way to collect data for the investigation. After piloting the questions on a single class, the questionnaire was administered to 555 MYP students (44% of whom responded) and 72 teachers (33% of whom responded) through the school's Moodle 2.9-based Virtual Learning Environment (VLE), which enabled the researchers to distribute the survey easily to both students and teachers. Both groups accessed the survey through the VLE and answered the questions online.

Although 80% of the students are non-native English speakers, English is the medium of instruction and approximately 90% of these second-language students are considered to be 'intermediate' English language learners or higher. With more than 40 nationalities at the school, it was not feasible to translate the survey questions into all of the students' mother tongues. Therefore, the survey questions were written in English and students were asked to respond in English, even though there was some concern that some questions may be misunderstood or misinterpreted in some way.

There were 11 statements based on our questions about rubrics given to both teachers and students and seven additional questions given to teachers. Respondents used a 1 to 5 scale (a.k.a. Likert scale), in response to a statement derived from the questions. Responses remained anonymous and confidential. Students were free to comment on rubric use in all school subjects and teachers answered according to their experience of using rubrics in their subject area. All respondents, both students and teachers, were asked to add their comments to supplement their answers and to clarify their concerns (Appendix A). Following the completion of the survey, the VLE software provided a basic analysis of the responses. The analysis gave an average of all the 1-5 responses and also showed the number of respondents for each choice from 1 to 5 on the scale and the percentage of all respondents the figure represents.

RESULTS

STUDENT SURVEY

The Likert scale used for the survey was atypical with 'strongly agree' requiring a response of '1' and 'strongly disagree' a response of '5'. For all of the questions, the mean student responses fell between option 2 ("agree") and option 3 ("neither agree nor disagree"). More precisely, the averages fell between 2.5 and 2.9, generally indicating tepid support for rubrics overall. A closer look at the data revealed that between 25 and 35 percent of respondents selected option 3 for each question and sixty-six entries (28%) in the comments section neither supported nor criticized rubrics. However, the remaining comments revealed a stark divide between those who found rubrics genuinely helpful and those who found them frustrating.

The most common frustration was not an unexpected one. Forty-three out of 237 respondents (18%) explicitly remarked that the rubrics were confusing or difficult to understand, not only making it hard for them to interpret their grade, but also how to improve. The second major frustration, reported by 47 students (20%) was a related one, namely that rubrics, whether used alone or in conjunction with other types of feedback, often failed to provide sufficient detail, either to adequately summarize student performance or to indicate paths toward improvement. Thus, rubrics failed many students due to a lack of actionable information.

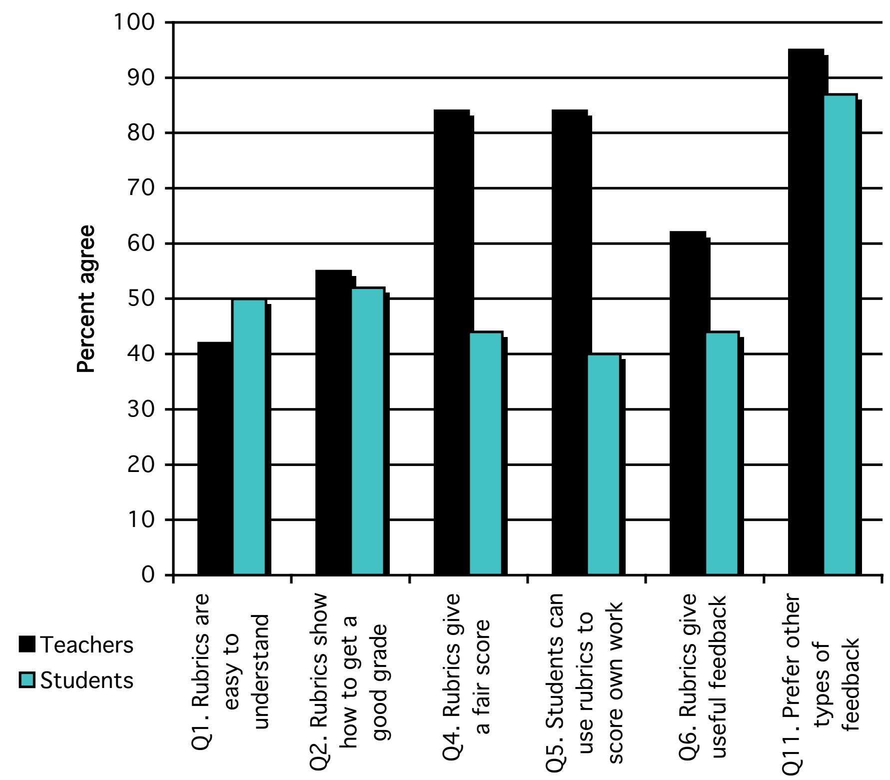


Figure 1. Comparison of teacher and student survey results. The "percent agree" axis corresponds to survey responses of '1' ("strongly agree") and '2' ("agree").

On the other hand, 52 comments (22%) were positive. These comments were not very detailed, though some specifically mentioned that rubrics were indeed clear and easy to use. Several positive comments were qualified, however. In fact, a few overlapped with the negative ones above. For example, three students liked the rubrics themselves but complained that some teachers did not interpret them as they were written.

TEACHER SURVEY

Teachers were generally more positive about the use of rubrics than the students were. 18 respondents (75%) recommended or strongly recommended rubrics for teachers or students while the remaining six (25%) were neutral. None of the teacher participants believed rubrics should not be used and a full 75% felt that rubrics accurately measured student performance. These figures suggest most teachers view rubrics as an effective tool for grading.

Teachers were less enthusiastic, however, when it came to statements such as, "Rubrics clearly tell students what they need to do to achieve a good grade" where agreement fell to 55%, or "The rubrics are easy for students for students to understand" which only 42% agreed with. Teachers also felt that students generally did not really use the rubrics; only 9 out of 24 (38%) believed that they did. While 63% felt that rubrics were easy to score, only 34% felt that "Rubrics are easy to construct". Fully 80% supported the statement that "Some rubrics are better than others", indicating that rubrics often failed to meet their promise of consistency. Finally, teachers overwhelmingly (95%) preferred other types of feedback. Teacher comments also reflected a much more qualified support of rubrics compared to the survey data. Teachers are certainly aware of the need for clarity on behalf of their second language students and several remarked that the IBMYP rubrics were not helpful in this regard. Others felt that the MYP rubrics were "narrow and rigid", on the one hand, or "too broad" on the other.

DISCUSSION AND CONCLUSION

This investigation started

with three main questions. The first was whether teachers and students viewed rubrics in the same way. A side-by-side analysis of the results of the teacher and student surveys (see Figure 1) reveals some visible parallels on one hand and some notable discrepancies on the other. For example, the survey data suggest close agreement between teachers (55%) and students (51%) on the question of whether IB rubrics clearly told students how to get a good grade (Q2). Conversely, there was a stark difference (teachers 84%, students 40%) over the question of whether students could use the rubrics to score their own work (Q5). These results show that teachers and students do not share the same views on all matters related to the use of rubrics.

The second research question concerned the fairness of grading with rubrics. The survey shows that students and teachers have strikingly different perspectives when it comes to the fairness of rubrics (Q4). While 84% of teachers felt that a rubric-derived score was fair, only 44% of students concurred, supporting Andrade's (2005) and Popham's (1997) findings that rubrics are often subjective. One student noted that, "I think that some teachers give the same comment on same [sic] rubric…even though our work is different." Confirming the veracity of this comment is beyond the scope of this research, but it does raise an important issue related to fairness of grading with rubrics: How can students be convinced that teachers use rubrics carefully and fairly? This question could be a line of inquiry for further research.

For a rubric to be fair, it would need to convey teacher expectations clearly to all students, but only 55% of teachers and 51% of students felt that the IB rubrics clearly told students how to get a good grade (Q2). Interestingly, it should be noted that although only 42% of teachers judged the IB rubrics easy for their students to understand (Q1), fully 84% felt that students could use the rubrics to score their own work (Q5). It is difficult to draw a conclusion from this seemingly contradictory piece of evidence. Perhaps teachers are expecting students to persevere in using the rubrics despite their lack of clarity.

The third research question concerned the use of rubrics for feedback. Since many teachers were supplementing IB rubrics with other types of feedback, the authors suspected that they did not feel that these rubrics are fully effective. In fact, the survey results confirm that both students and teachers find that the capacity of the IB rubrics for giving detailed and useful feedback is indeed limited. Comments by students about the lack of detail the rubrics give for feedback and the confusing nature of the language used in the rubrics support Popham's (1997) findings that rubrics are unclear and open to interpretation. One student commented that he found the rubrics to be vague and confusing. On the other hand, written teacher comments gave him a clear picture not only of ways he could improve, but also of what he did correctly. Comments like these cast doubt on the ultimate usefulness of the IB rubrics for giving feedback. In fact, while only 8% of teachers felt that rubrics did not give useful feedback (Q6), 24% of students felt this way. These survey results suggest that many students do not find rubrics as suitable for this purpose as their teachers do.

Most telling, perhaps, was the broad consensus about how rubrics compared to other types of feedback (Q11). Both teachers (95%) and students (86%) overwhelmingly preferred either teacher-written comments or an oral feedback session to either rubrics alone or highlighted rubrics. Of course, teachers are often constrained by time, but both groups recognize the need for in-depth feedback, which seems to be beyond the scope of the IB rubrics used alone. This suggests that IB rubrics should always be used in conjunction with other types of feedback.

As mentioned above, rubrics have been promoted on the basis of certain advantages: they are meant to standardize grading and to be easy for teachers to use. Rubrics are also meant to encourage students to evaluate themselves over the course of the unit, thereby directing their focus to improvement. However, less than half (42%) of the teachers believed the rubrics were easy for their students to understand and teacher comments such as 'narrow and rigid' and 'too broad' suggest a lack of clarity in the rubrics. Therefore, although the survey results indicated that IB rubrics were easy for teachers to implement, given the fact that many teachers and students found them unclear, the IB rubrics failed to achieve the goal of grade standardization. In addition, as the IB rubrics are used in a summative (end-of-unit), rather than in a formative (mid-unit) capacity, they did not particularly help students adjust their performance over the course of the unit.

Rubrics have been criticized as invalid, unreliable, unfair, and/or subjective to some degree (Andrade, 2005; Popham, 1997). Our research did not aim to fully evaluate the reliability and validity of the IB rubrics. However, the aggregate student response revealed serious doubts regarding the fairness and objectivity of the use of IB rubrics. A closer look at the rubrics themselves would certainly be useful in terms of evaluating their potential. Further investigation should attempt to determine whether there are specific qualities of rubrics that students and teachers consider more effective or whether rubrics have inherent limitations that cannot be overcome.

In conclusion, the study revealed both agreement and disagreement between teachers and students concerning the use of IB rubrics for grading and feedback at the authors' school. The difference in teacher and student views about the fairness of rubric-scored grades could be of major importance to the wider educational community. Similarly, the difference in teacher and student views about the use of rubrics for giving feedback should be a cause for concern. The study suggests that the use of IB rubrics for both grading and giving feedback should be re-evaluated and steps taken to address the students' concerns so that rubrics are not used for feedback in isolation. Rather, face-to-face interaction and detailed written feedback should also be part of the assessment equation. Other institutions in which rubrics are used, whether IB or not, may also wish to consider the relevance of these research findings within their own context.

APPENDIX A

This appendix contains the teacher and student survey statements and the questions from which they were derived. The 11 overlapping statements common to both surveys were:

1. The rubrics are easy to understand.
2. Rubrics clearly tell what is needed for a good grade.
3. During assessment, students use rubrics to help themselves.
4. Rubrics give students a fair score.
5. Students can use rubrics to score their own work.
6. Rubrics provide useful feedback.
7. Rubrics help students to learn.
8. Some rubrics are better than others.
9. Teachers use simplified rubrics (for the benefit of ESL students).
10. Teachers use a variety of assessment methods (besides rubrics).
11. What is the best type of feedback after an assessment?
    a. rubric only
    b. highlighted rubric
    c. teacher written comments
    d. teacher oral comments

The following were the seven additional questions given to the teachers:

1. Which levels of English did they teach? (Respondents could choose up to 4 levels out of 8 options ranging from beginner level 2 to native speaker fluency.)
2. How often do you share rubrics with students after an assessment?
3. How often do you share rubrics at the beginning of a unit?
4. Are rubrics easy to construct?
5. Are rubrics easy to score?
6. Do rubrics give me an accurate picture of student performance?
7. Do you recommend rubrics for teachers and students?

The questions the statements were derived from:

1. Are rubrics used easy to understand?
2. Do the rubrics clearly tell what is needed for a good grade?
3. During assessment, do students use rubrics to help themselves?
4. Do rubrics provide a fair score?
5. Do students use rubrics to score their own work?
6. Do rubrics provide useful feedback?
7. Do rubrics aid learning?
8. Are some rubrics better than others?
9. Do teachers use simplified rubrics?
10. Do teachers use a variety of assessment methods?
11. What is the best type of feedback after an assessment?
    a. rubric only
    b. highlighted rubric
    c. teacher written comments
    d. teacher oral comments

REFERENCES

Andrade, H. G. (2005). Teaching with rubrics: The Good, the bad, and the ugly. College Teaching, 53(1), 27-30. doi:10.3200/CTCH.53.1.27-31

Andrade, H. (2007). Self-assessment through rubrics. Educational Leadership, 65(4), 60.

Brookhart, S. M. (2008). How to give effective feedback to your students. Alexandria, VA: American Society for Curriculum Development.

Brookhart, S. M. (2013). Develop a student-centered mindset for formative assessment. Voices from the Middle, 21(2), 21-25. Retrieved from http://search.proquest.com/docview/1464751036?accountid=14187

Center for Advanced R esearch in Language Acquisition. (2015). Types of rubrics. Retrieved from http://www.carla.umn.edu/assessment/vac/improvement/p_5.html

Cooper, B. S., & Gargan, A. (2009). Rubrics in education: Old term, new meanings. Phi Delta Kappan, 91(1), 54-55. doi: 10.1177/003172170909100109

DePaul University Teaching Commons. (2015). Types of rubrics. DePaul University. Retrieved from http://resources.depaul.edu/teaching-commons/teaching-guides/feedback-grading/rubrics/pages/types-of-rubrics.aspx

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. Educational Research Review, 2, 130-144. doi:10.1016/j.edurev.2007.05.002

Kohn, A. (2011). Feel-bad education and other contrarian essays on children and schooling. Boston: Beacon Press.

Mueller, J. (2014). What is authentic assessment? Authentic Assessment Toolbox. North Central College. Retrieved from http://jfmueller.faculty.noctrl.edu/toolbox/whatisit.htm

Newkirk, T. (2000). A mania for rubrics. Education Week, 20(2), 41.

Popham, W. J. (1997). What's wrong and what's right with rubrics. Educational Leadership, 55(2), 72-75. doi: 10.1111/j.1745-3992.1993.tb00517.x

Valenza, J. K. (2000). Students and teachers alike can benefit from rubrics: Selections from an interview with Dr. Andi Stix. Retrieved from http://www.uwstout.edu/soe/profdev/resources/upload/Students_and_Teachers_Alike_Can_Benefit_from_Rubrics.pdf

Wiggins, G. (1993). Assessing student performance: Exploring the purpose and limits of testing. San Francisco: Jossey-Bass, Inc.

Author Biographies

Douglas Glenn was the Secondary Librarian at Suzhou Singapore International School from 2011 to 2015. He is currently a consultant living in Panama. his main professional interest is helping secondary students master academic research.

Email:
dglenno47@gmail.com

Graham Morton currently teaches English Language Acquisition at SSIS. He has an MSc in TESOL and has taught at international schools inTaiwan, Vietnam and China. His professional interests include vocabulary development, and contrastive rhetoric for international school students.

Email:
grahammorton@ssis-suzhou.net

IN-HOUSE RATING SCALE DEVELOPMENT

Simon Dawson

ABSTRACT

The rating scale is a central part of the assessment of spoken and written performance. Publicly available scales often do not fit the needs of specific assessments and therefore a unique scale is often created from scratch. The development of a valid and reliable rating scale is time-consuming, especially if it is done without experience or guidance. This paper attempts to provide guidance to practitioners on rating scale development. To do this, the paper gives an account of an attempt to revise a rating scale for use in the assessment of oral presentations. Using action research, the study focuses on the lessons that were learnt and ends with a set of recommendations.

摘要

等级量表是评估口语和写作表现的核心组成部分。然而，可获取的公开量表往往不能满足特定评估的需要，因此，经常需要从头建立量表。 开发一个有效且可靠的定级量表很耗时，尤其是在没有经验或指导的情况下。本文旨在为等级量表开发者提供指导。基于此目的，文章详述了一个针对口头报告测试而修改一个定级量表的尝试。使用行动研究方法，本研究聚焦所取得的经验和教训，并在结尾部分提出了一组建议。

INTRODUCTION

A sound rating scale is an essential part of a reliable and valid assessment of spoken and written language; it clearly sets out the construct being measured, provides accurate descriptions of the typical performance expected at different levels, and guides raters in making consistent judgements that are in line with an agreed standard. For these reasons a great deal of time and money is invested in the development of rating scales (e.g. Galaczi, ffrench, Hubbard, & Green, 2011; Fulcher, 2003; North, 2003; Turner, 2012). Some larger organisations make their scales available to the public (e.g. IELTS and TOEFL) and so these can be used in institutions without the resources to develop their own. However, a rating scale is designed for a specific purpose and the more generic, publicly available scales often do not suit the needs of specific educational settings. For this reason, institutions will often develop rating scales in-house, often with limited time and resources. While there is some useful guidance for practitioners regarding scale development, for example, the Council of Europe (2011) publication on the development of the Common European Framework of Reference (CEFR), this type of guidance is limited. This paper, then, attempts to provide guidance to practitioners who find themselves faced with the task of creating or developing a rating scale with limited resources. To do this, the researcher has taken an action research approach to gain first-hand insight into in-house rating scale development. The development of a rating scale for the purpose of assessing an EAP speaking exam is described first, and then recommendations based on the experience are presented. The research setting is the Centre for English Language Education (CELE), a department of the University of Nottingham, Ningbo, a Sino-British university in China. CELE has a large EAP program of approximately 1,500 pre-undergraduate students and 170 pre-masters students.

BACKGROUND INFORMATION: RATING SCALE DEVELOPMENT METHODOLOGIES

The literature on rating scale development methodologies outlines two main approaches: the armchair approach and the empirical approach (Lim & Galaczi, 2013). In the armchair approach, a scale is created based on an expected range of performance and draws solely on expert knowledge. Also called the intuitive method (Council of Europe, 2011, p.208) and the a priori method (Fulcher, Davidson, & Kemp, 2010), this method has the advantage of being based on theoretical views about the development of second-language (L2) ability or from learning objectives set out in a course curriculum. The problems with this approach stem largely from the fact that expectations often do not match reality. In contrast, the empirical approach, also known as performance data-based methods, (Fulcher, Davidson, & Kemp, 2010) is data-driven with evidence guiding the creation of the scale and the writing of descriptors. Qualitative methods (workshops, observation) as well as quantitative methods, for example, discriminant analysis and item response theory (Council of Europe, 2011, p. 210), are used to develop the assessment instrument. The big advantage with the empirical approach is that the scale not only reflects actual learner performance but also the way they are referred to by assessors.

REPORTING OF RESEARCH ACTIVITY: REVISION OF RATING SCALE FOR ACADEMIC ORAL PRESENTATIONS

The project used in this paper is the revision of a rating scale used for the assessment of academic oral presentations (AOP). The scale is relevant to several users: students to know how they will be assessed and what their scores mean; tutors in preparing students for the assessment; assessors to guide decision making during the assessment; and trainers who use the scale in the training of tutors to carry out assessments of AOPs. With such wide use, the document plays an important role in a course that is compulsory for over 1,500 students. Because of the importance of the document, it is under close scrutiny and therefore under constant review. In its latest review, two areas in need of revision were identified: 1. the level of detail of the descriptors, and 2. the number of components to score.

Looking firstly at the level of detail of the descriptors, the scale descriptors were designed to be sparse for ease of use. In the scale descriptors, the only real change across bands is the adjective used to describe the particular features of the category. For example, in the category of Cohesion (Structure and Linking Points), the differentiation between the five levels of organization is made simply by changing the adjective, i.e. Effective organization changes to Satisfactory organization. Such sparse description does not give users sufficient information about the performance expected at different levels across the scale and there was call from tutors in particular for more detail in descriptions.

The second aspect identified as in need of change was the number of components to score. The assessment required assessors to provide eleven component scores. As reported by assessors and experienced first-hand by the researcher, assigning such a high number of scores with any degree of consistency or accuracy during a live assessment was extremely challenging and often simply not possible. The aims of the revision then were to 1. expand the descriptions of performance to provide better guidance, and 2. reduce the number of components to be scored to reduce the load on assessors.

OVERVIEW OF THE SCALE DEVELOPMENT ACTIVITIES THAT TOOK PLACE

STEP 1: CREATING A DRAFT SCALE

AIM 1: REDUCE THE NUMBER OF COMPONENTS TO BE SCORED

To reduce the number of components to be scored would require either dropping components or merging them. As a summative assessment, the AOP construct is basically the course learning outcomes: a list of the things students should be capable of by the end of the course. The person responsible for the course required that all the components were retained in the assessment so none could be dropped; that left the second option which was to merge components. To do this the researcher went back to the construct to see if it could be reorganised. Banerjee and Wall (2006) outline a procedure that can be used to establish a construct. To begin with, a list of all the relevant features that need to be incorporated into the assessment is compiled, then the items are organised according to natural groups. Following this idea, the researcher set up a focus group with EAP tutors from the department who taught and assessed on the AOP course, and were therefore familiar with the terms used in the course learning outcomes. As in Banerjee and Wall's (2006) study, a list of the relevant features to be incorporated in the assessment was compiled and then physically cut into pieces. Participants were asked to group the pieces in any way they felt was intuitive. The three tutor pairs all came up with different groupings, which was interesting as it indicated that there is no single correct way to group the statements, as the researcher had hoped there would be. Nevertheless, the groupings created and the justifications given during the session helped to inform a revised construct framework which was considered to better represent the construct and had a slightly reduced number of scorable components.

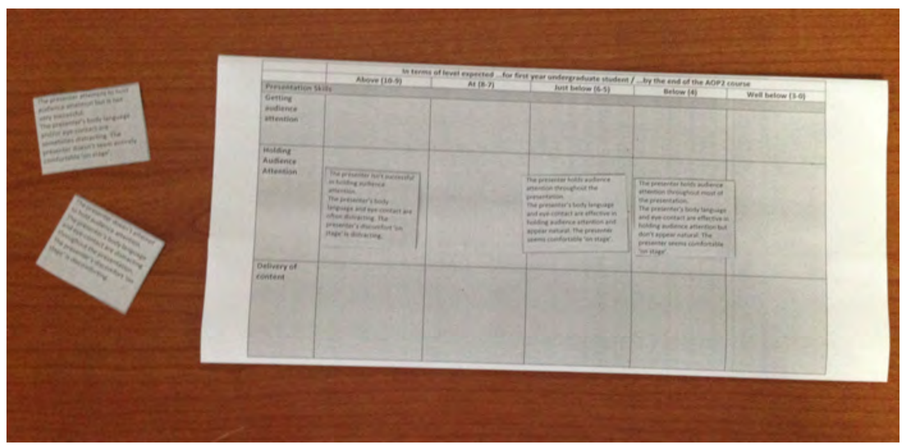


Figure 1. Uncovering problem areas.

AIM 2: EXPAND PERFORMANCE DESCRIPTIONS

With the components set, the next issue to address was the descriptors. Due to limited time, the arm-chair approach was adopted and a set of descriptors was written by the researcher using knowledge from teaching the AOP course and assessing AOP performances. In developing the descriptors, the researcher discovered the difficulty of attaining consistent labelling throughout and keeping wording brief whilst also clear for a user.

STEP 2: UNCOVERING PROBLEMS IN THE DESCRIPTORS

With a draft set of scale descriptors written, the next stage taken by Galaczi et al. (2011) was to uncover problem areas. They did this using both quantitative and qualitative methods. Rasch analysis was used in the quantitative study to establish which descriptors were consistently applied to which performances. Due to lack of experience, we decided to opt for a qualitative approach: a focus group. Participants were asked to re-order the descriptors which had been cut into separate pieces of paper. In preparation for the session, the scale was divided and a blank grid created with the scoring criteria on the y-axis and a 5-point scale on the x-axis. The individual descriptors were distributed to tutors one set at a time and tutors were asked to put them along the scale at the point they felt they best represented. To make the task slightly more challenging, and therefore force participants to think carefully about where to place the descriptors, participants were not given a full set of descriptors so they also had to identify where the gap created by the missing descriptors should be placed (Figure 1). The premise was if participants were able to re-order them in the same way they are ordered on the master scale, then this would support the descriptors. Indeed, problem descriptors were identified quickly by participants as they were the ones which participants were unable to order easily. The two main problems were lack of clear differentiation from adjacent descriptors and lack of clarity in wording. This method was an efficient way to have tutors familiar with the assessment look closely at the draft descriptors and be forced to make decisions based on their appropriate placing on a scale. After the session, the problem descriptions were revised and a second draft scale was produced.

In their second study, Galaczi et al. (2011) used a verbal protocol to see whether (and how) assessors refer to the descriptors during an assessment. In the present study, a similar method was used with EAP tutors attending a trialling of the second draft of the scale. Four participants familiar with assessing AOPs watched two video presentations and used the descriptors to rate the videos. Participants were provided with note paper to record comments as they completed the task. In this session, rather than giving participants the full 8-component scale (this was seen as unrealistic as the participants were unfamiliar with the new descriptors), they were given descriptors for only three components to score. The researcher clarified meaning of the annotations while collecting them and then used these to inform the creation of a full scale ready for trial.

STEP 3: TRIALLING THE FULL SCALE TO SEE IF RATERS ARE ALIGNED

Similar to the second study run by Galaczi et al. (2011), this study sought to see to what extent raters were in agreement when using the scale to rate an AOP video performance. A secondary aim was to gather further feedback on how closely the descriptors reflect actual performances.

Eight departmental EAP tutors used the trial scale to rate two video oral presentations. Participants independently registered their ratings by highlighting statements across the categories they felt best matched the performance they saw. It was found that raters showed consistent agreement in their evaluation, which suggests the rating scale works well. Raters also reported the scale was an improvement on what they had previously used with the descriptors providing more concrete guidance to decision making. Moreover, it was commented that for any new scale, a period of familiarization and moderation is needed for assessors to begin using the tool as it is designed to be used.

RECOMMENDATIONS RESULTING FROM THE RESEARCH

While the outcome of the scale development was useful for the purposes of this paper (i.e. a revised scale for use in the large-scale assessment of academic oral presentations), of greater importance are the recommendations resulting from the experience, as follows:

1. WORK FROM ACTUAL PERFORMANCES

In the same way that Galaczi et al. (2011) began the development process with experienced assessors identifying areas in need of revision and a committee setting the parameters, so did this study. The research committee agreed that the number of components should be reduced and the descriptions expanded. In hindsight, the importance of reducing the number of components to fewer than five, as recommended by Luoma (2004), was apparent. The scale development should have begun not with a discussion of changes that need to be made but with a viewing of the task being performed (i.e. an academic presentation being given). This would help to ensure that the assessment ▷

	10	8	6	4	2	0
Eye contact	x	xxxx	x			
Use of visuals		x	xxxxx	x	x	
...						

Figure 2. Example of scoring using Distributional Assessment (DA) Model

fits the reality, rather than vice versa, which is a point Alderson (1991) makes. The committee, along with experienced assessors, should watch the performance and take notes of features that stand out. These features, with reference to the course learning objectives, should be compiled to make up the components to be scored. In this way, it would have been established early on that it is not possible for an assessor to successfully handle more than five components during a 10-minute presentation. As well as the number of components to score, the number of bands in the scale (5) was set to fit neatly into the university scoring system of a 0 – 100 scale. In fact, for some components, there clearly were not five sufficiently differentiated levels of performance (e.g. presentation opening or referencing). This again supports the importance of working from examples of performance early on so that only true distinctions are incorporated, rather than forced ones.

Similarly, in the writing of descriptors, working from actual performances with tutors would have helped capture better the true points that distinguish presentations from one another and the meta-language used by assessors when evaluating performance. A method to do this is described in the Council of Europe (2011) document in which workshop participants rank performances, explaining their ranking. This method captures the most salient features used by assessors used to differentiate between levels as well as the language used by the raters to describe performances. This would help to ensure distinctive differences are included in the descriptors, which in turn will help guide assessors better in their assessment judgements.

2. UNDERSTAND HOW ASSESSORS USE A MARKING SCALE AND LET THAT INFORM THE SCALE DESIGN

Considering the fact that a scale is a tool used for guiding decisions during assessment (and for spoken performance, the assessment is usually live), then the way it is used in practice should be considered when designing the scale. From personal experience of live assessment, a tally system in which scores for each component are continuously awarded during the performance is a practical form of record keeping when there are numerous components (the overall score for the component is then made by averaging the scores at the end). Kane (1986) put forward a similar approach for assessment of performance: the Distributional Assessment (DA) Model. With this model, rather than having the assessor observe a performance and try to simultaneously retain global impressions of several independent criteria, the assessor records every judgment they make as they witness the relevant behavior (figure 2).

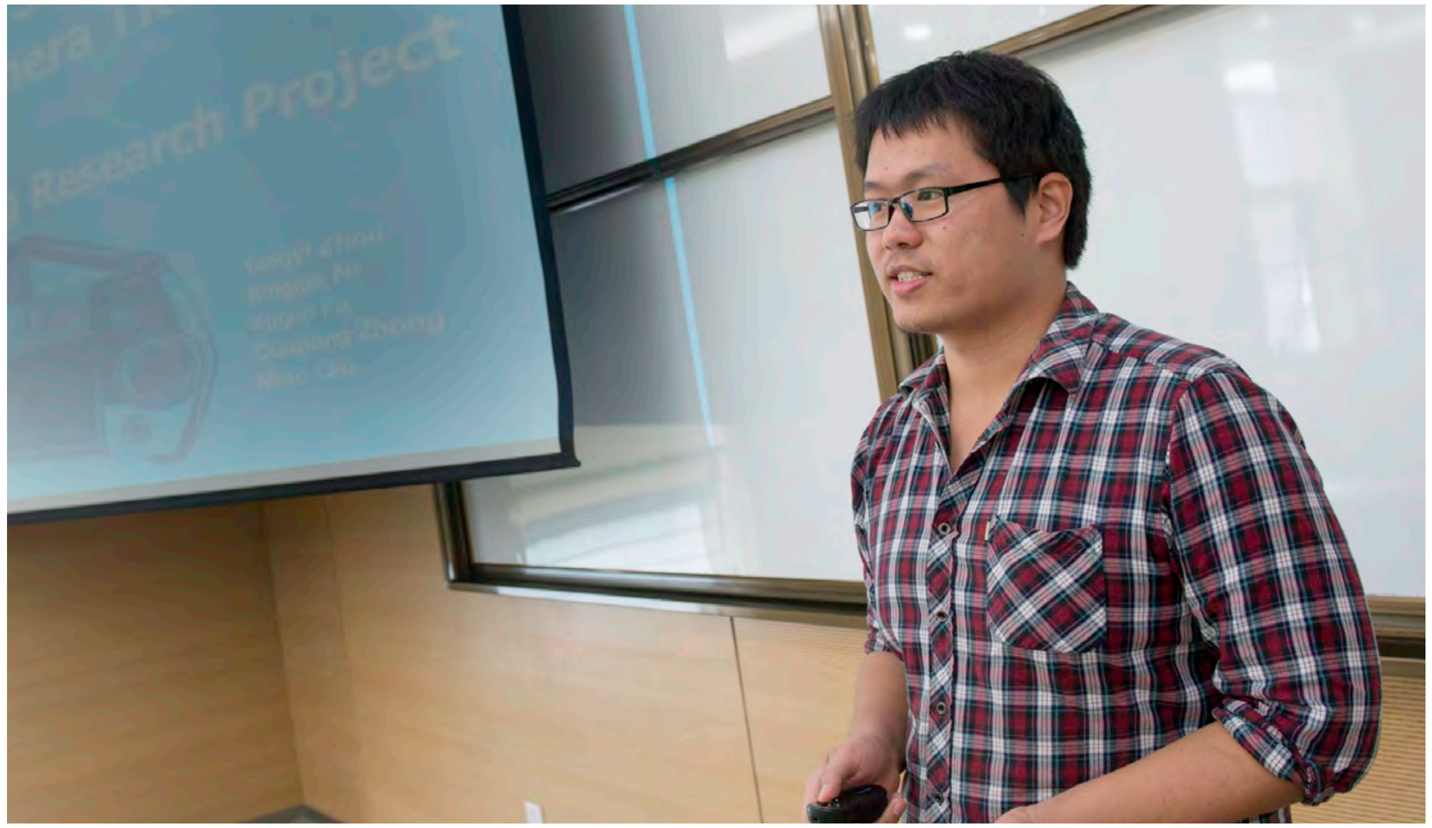
It is therefore worth investigating how assessors approach the task without any guidance and then using that understanding in the assessment tool design.

3. MAKE USE OF WHAT IS ALREADY THERE

In point 1 above (Work from actual performances), it is suggested that language actually used by assessors should be employed as much as possible in scale development. Jeffrey (2015) looked to assessor comments made in coursework feedback to build descriptors for a writing marking scale. Common features assessors used to distinguish adjacent scores were identified which meant assessor meta-language was captured and there was no need to bring tutors in specifically to score and add comments to scripts. Further, the comments were made in regular assessment activities and so highly authentic. While the current assessment (AOP) does not have such a record of assessor comments, as written feedback is not given, there is a record of tutor comments given to benchmarked AOP videos, which are used for assessor training.

CONCLUSION

This study set out to develop a reliable rating scale for the assessment of oral presentations. Through this experience, several lessons were learnt including the importance of a clear understanding of the test construct, the importance of not making the descriptors too contrived (they should reflect reality, not try to dictate it) and the importance of considering the user. Overall, with well-written, empirically-grounded descriptors the task of making a fair assessment becomes easier. A rating scale that is not representative of actual performance, that is unclear or that is difficult to use during an assessment is unsuitable for use and, apart from leading to potentially unreliable or invalid assessment judgements, can create a lack of confidence in users. To make a scale that meets the needs of users requires craftsmanship and sound judgement, but the time spent in getting it right should pay off. In the end, while it was not possible to demonstrate that the rating scale created in this study was producing consistent results, the process of developing the scale has produced a piece of work that will hopefully help to guide others in rating scale development projects of their own.



REFERENCES

Alderson, J.C. (1991). Bands and scores. In J. C. Alderson & B. North (Eds.) Language Testing in the 1990s: The Communicative Legacy (pp 71-86). London: Macmillan.

Banerjee, J. & Wall, D. (2006). Assessing and reporting performances on pre-sessional EAP courses: Developing a final assessment checklist and investigating its validity. Journal of English for Academic Purposes, 5, 50-69. doi:10.1016/j.jeap.2005.11.003

Council of Europe (2011). Common European Framework of Reference for Learning, Teaching, Assessment. Council of Europe. Retrieved from http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

Fulcher, G. (2003). Testing second language speaking. London: Longman/Pearson.

Fulcher, G., Davidson, F., & Kemp, J. (2010). Effective rating scale development for speaking tests: Performance decisions trees. Language Testing, 28(1), 5–29. doi:10.1177/0265532209359514

Galaczi, E., ffrench, A., Hubbard, C. & Green, A. (2011). Developing assessment scales for large-scale speaking tests: a multiple-method approach. Assessment in Education: Principles, Policy and Practice, 18(3), 217-237.

Jeffrey, R. (2015). Using feedback comments to develop a rating scale for a written coursework assessment. Journal of English for Academic Purposes, 18, 51 – 63. doi:10.1016/j.jeap.2015.03.002

Kane, J. S. (1986). Performance distribution assessment. In R. A. Berk (Ed.), Performance Assessment: Methods and Applications (pp. 237-273). Baltimore: Johns Hopkins University Press.

Lim, G. & Galaczi, E. (2013). Pre-conference workshop: Rating scales and raters in speaking assessment. Language Testing Research Colloquium 2013, Seoul, South Korea.

Luoma, S. (2004). Assessing Speaking. Cambridge: Cambridge University Press.

North, B. (2003). Scales for rating language performance: Descriptive models, formulation styles, and presentation formats. TOEFL Monograph, 24.

Turner, C. E. 2012. Rating scales for language tests. The Encyclopedia of Applied Linguistics. doi: 10.1002/9781405198431.wbeal1045

Author Biography

Currently EAP Tutor at University of Nottingham in Ningbo; soon to be EAP tutor at Bristol University. I teach EAP courses and design and develop EAP assessments. Recent assessment projects include the revision of rating scales for seminar discussion skills and interview speaking skills.

Email:
simon.dawson@nottingham.edu.cn

INSITES

This section highlights some useful learning and teaching websites that can help with planning, teaching and professional development. This time we look at assessment-related resources.

Jackie Hemingway & Sarah Butler

LANGUAGE TESTING RESOURCES WEBSITE

languagetesting.info

The Language Testing Resources website is a reference guide to resources related to language assessment. The site is designed and maintained by Glenn Fulcher (University of Leicester), who is widely published in the field, and is co-editor of the journal Language Testing. A range of resources covering a wide variety of language assessment related topics is provided: links, articles, a bookstore, videos, podcasts, feature topics, scenarios and statistics.

Launched in 1995, the site continues to be regularly updated. On the homepage, news feeds are updated every four hours with worldwide language testing news. The "most visited websites" in assessment are displayed, based on results from major search engines. A list of upcoming conferences also enables users to stay up to date with the latest research. The job listings section, in cooperation with the International Language Testing Association (ILTA), is updated regularly with opportunities in the field of testing and assessment. At the time of this review, about 30 jobs were featured, mainly in the US but also including opportunities in China.

The "articles" and "links" sections are well-organised and extensive. All the articles are from a range of sources including over thirty respected language and teaching journals, and are freely available and hyperlinked. At the bottom of the articles page, a list of online collections and journals covering assessment issues may also be helpful for researchers. The links page includes language testing associations and providers, specific tests, models, frameworks, statistics and online tests. Further reading is available on the bookstore page, with books available for purchase via Amazon categorized by topic.

Users can also watch recent videos made by speakers from various institutions via YouTube (requiring access to a VPN within China). Additionally, twenty videos have been produced specifically for the site. These are mostly less than ten minutes long, and although looking rather outdated and with poor video quality in the early versions, do provide a brief introduction to assessment issues such as validity and reliability, and to specific assessment types such as integrated assessment. These materials are designed as an introduction for students of language testing, and may be too general for experienced practitioners in the field.
More in-depth discussions are available through Language Testing Bytes, a podcast series from the journal Language Testing and available on the site, as well as on iTunes. Programmes are released 3 - 4 times a year, although from 2016, they will be reduced to biannually. The July 2015 podcast discusses diagnostic language testing.

The "features" and "scenarios" sections are designed for classroom use with students of language testing to encourage consideration of particular testing situations. For example, the High stakes testing feature provides a video and discussion questions about the high pressure Gaokao examination in China.

Of more interest to those working in assessment may be Excel spreadsheets for classical test analysis; five spreadsheets are available for download, including Cronbach's alpha, a statistical formula for calculating test reliability.

Although the site design could be improved to be less cluttered, the menus and search facility make the site generally easy to navigate. Information alert services also allow users to stay up-to-date with language testing issues by subscribing to feeds on news, articles, YouTube videos, podcasts and job postings. The breadth of materials provided means much can be found relevant to both ELT and EAP. This is a useful resource guide for both those new to language testing and for more experienced test designers and assessors.

SEMINARS FOCUSING ON ASSESSMENT

englishagenda.britishcouncil.org/seminars

The British Council's Continuing Professional Development website has a number of seminars, some of which focus on assessment. Many seminars are subdivided into shorter bite-sized recordings which are easy to access and digest.

Some of the seminar recordings contain "session notes" which give some background to the speaker and content. There may also be tasks for the viewer to complete before watching, or follow-up tasks, and there may be the option of joining a discussion forum. Print versions of the training session and links to appropriate articles, supplementary information and further reading are all useful additions. The seminars are filmed in such a way that they focus on the presenter, but cutaways to the PowerPoint slides help the viewer follow the presentation more easily.

Many of the seminars which focus on assessment and testing deal specifically with IELTS testing. A good example is that by Sam McCarter, an author and presenter. One of his listed presentations, Teaching IELTS: Producing materials for the academic reading and writing modules, recommends strategies for teaching IELTS and takes a functional approach to IELTS teaching material. His presentation is clear, easy to follow and offers a number of practical ideas that can be implemented immediately. An example is students paraphrasing exam questions in order to get greater understanding of the task. Overall his approach focuses on student awareness and self-evaluation of learning.

Other presentations focusing on testing include one by Fiona Aish and Jo Tomlinson on the topic of preparing lower-level students for IELTS. This is well-resourced and includes slides and a hand-out.

Barry O'Sullivan gives an overview of the development of assessment and language testing from its origins up to the present day and looks at how it may develop in the future. His presentation is entertaining and gives an insight into how tasks in exams were developed, particularly within the context of the teaching approaches popular at the time.

Overall, the training sessions give an insight into various aspects of testing and possible approaches when teaching. While the seminars mostly focus on IELTS preparation, they offer ideas of how to maximise learning which may be used in EAP or EFL classes. There is a good balance of theory and practice with a variety of topics and presenters. The website is well-designed and easy to access and navigate. The side bar containing details of upcoming seminars is a useful addition. The inclusion of preparation tasks, an online discussion forum, and further resources gives the user a great deal of choice on this website and is well worth investigating further by individuals and can be incorporated into an institution's continuing professional development programme.

Author Biographies

Jackie Hemingway has taught EAP at XJTLU since September 2011. Prior to joining XJTLU she taught General and Business English in several countries including Austria, Spain, Japan and Oman.

Email:
jacqueline.hemingway@xjtlu.edu.cn

Sarah Butler is Associate Director of Pre-sessional English courses at CU Services Coventry University. Her areas of interest include teacher development and learner autonomy.

Email:
ab6631@coventry.ac.uk

UPCOMING CONFERENCES

Although every effort has been made to supply accurate information, readers should visit the conference websites to receive the latest updates. Many conferences choose to extend the date for abstract proposals to ensure the maximum number of presenters.

Seth Hartigan

CONFERENCES WITH OPEN PROPOSAL DEADLINES

8TH INTERNATIONAL CONFERENCE ON TESOL: BRIDGING ELT THEORY AND PRACTICE: INNOVATIONS AND RESERVATIONS IN THE DIGITAL AGE

April 15-17, 2016, English Language Centre (ELC), Shantou University, Guangdong, China

Proposal due: November 30, 2015

Website: elc.stu.edu.cn/conf2016

THE ASIAN CONFERENCE ON LANGUAGE LEARNING 2016: ACLL2016

April 28 - May 1, 2016, Art Center of Kobe, Kobe, Japan

Proposal due: January 1, 2016

Web: iafor.org/acll2016-call-for-papers

13TH EUROPEAN ASSOCIATION FOR LANGUAGE TESTING AND ASSESSMENT (EALTA) 2016 CONFERENCE

May 5-8, 2016, La Universidad Politécnica de Valencia, Valencia, Spain

Proposal due: November 10, 2015

Website: www.ealta2016.com

CENTRE FOR ENGLISH LANGUAGE COMMUNICATION (CELC) SYMPOSIUM 2016: STRENGTHENING CONNECTIVITIES IN ELT: PEDAGOGIES, DISCIPLINES, CULTURES

May 25-27, 2016, Stephen Riady Centre, University Town, National University of Singapore

Proposal due: November 15, 2015

Website: nus.edu.sg/celc/symposium

LANGUAGE CULTURE AND MIND 7: SIGNS OF LIFE: CULTURAL CONTACT, CHANGE AND CONTINUITY IN LANGUAGE, THOUGHT AND IDENTITY

June 1-4, 2016, Hunan University, Changsha, China

Proposals: open

Website: languageculturemind.org

JAPANESE ASSOCIATION FOR LANGUAGE TEACHING (JALT) TASK-BASED LEARNING SIG: 3RD TASK-BASED LANGUAGE TEACHING IN ASIA CONFERENCE

June 25-26, 2016, Ryukoku University, Kyoto, Japan

Proposal due: March 15, 2016

Website: www.tblsig.org/conference/cfp

THE EUROPEAN CONFERENCE ON LANGUAGE LEARNING 2016: ECLL 2016

June 29 - July 3, 2016, The Waterfront Hotel, Brighton, England, United Kingdom

Proposal due: March 1, 2016

Website: www.kate.or.kr

THE BRITISH ASSOCIATION FOR APPLIED LINGUISTICS BAAL 2016 CONFERENCE

September 1-3, 2016, Anglia Ruskin University, England, United Kingdom

Proposal due: TBD

Website: www.baal.org.uk/baal_conf.html

JAPAN ASSOCIATION FOR LANGUAGE TEACHING'S (JALT) 2016: 42ND ANNUAL INTERNATIONAL CONFERENCE ON LANGUAGE TEACHING AND LEARNING

November 25-28, 2016, Aichi Industry & Labor Center – WINC Aichi, Nagoya, Aichi Prefecture, Japan

Proposal due: TBD

Website: jalt.org/main/conferences

CONFERENCES WITH CLOSED PROPOSAL DEADLINES

INTERNATIONAL LANGUAGE TESTING ASSOCIATION (ILTA) LANGUAGE TESTING RESEARCH COLLOQUIUM (LTRC) 2016 PALERMO CONFERENCE: LANGUAGE CONSTRUCTS, CONTEXTS, AND CONTENT IN CLASSROOM AND LARGE-SCALE ASSESSMENTS

June 20-24, 2016, University of Palermo, Sicily, Italy

Website: www.iltaonline.com/LTRC/index.php

2016 INTERNATIONAL WRITING ACROSS THE CURRICULUM CONFERENCE: WRITING ACROSS DIFFERENCE

June 23-26, 2016, University of Michigan, Ann Arbor, Michigan, USA

Website: iwac2016.org

3RD INTERNATIONAL CONFERE ON ACADEMIC WRITING: BUILDI BRIDGES THROUGH ACADEMIC WRITING: RESEARCH, POLICY, A PRACTICE

June 27-28, 2016, Mofet Institute Tel Aviv, Israel

Website: conference2016.macam ac.il/

Author Biography

Seth Hartigan has taught Academic English for the last ten years at XJTLU, Renmin and Tsinghua Universities. His academic interests include the philosophy of education and the science of learning.

Email:
seth.hartigan@xjtlu.edu.cn

Xi’an Jiaotong-Liverpool University
西交利物浦大學

CONTACT

English Teaching in China
etic@xjtlu.edu.cn
www.xjtlu.edu.cn