

Improving the **Quality** and **Consistency** of **EAP** **Assessment** through **Inter-University Collaboration**

Sam Barclay

Nottingham Trent University



“language centres [need to] have a dedicated team of colleagues with timetabled commitments to creating, editing and standardising the marking of...all forms of testing taking place. Unfortunately,...there is limited evidence of this in place.” (BALEAP mail list contributor, 2017)

“[I hope to promote] collaborative research and learning opportunities, linked to EAP assessment Literacy, both within and beyond individual institutions. Such opportunities could include cross-institutional groups or pairings...” (Manning, 2013).

Outline

1. Background
2. Survey data on assessment practices in EAP
3. Example of collaboration and assessment development
4. Reflections on importance of collaboration
5. Call for more inter-university collaboration

Background

- Multifaceted role of pre-sessional EAP courses (Seviour, 2015):
 - foster language and skills required to successfully participate in an academic context.
 - summatively assess knowledge to determine student readiness.
- Need for EAP professionals to develop language assessment literacy (LAL) (Manning, 2013; Schmitt & Hamp-Lyons, 2015) and for greater dissemination of assessment practices (Manning, 2013; Schmitt, 2017).

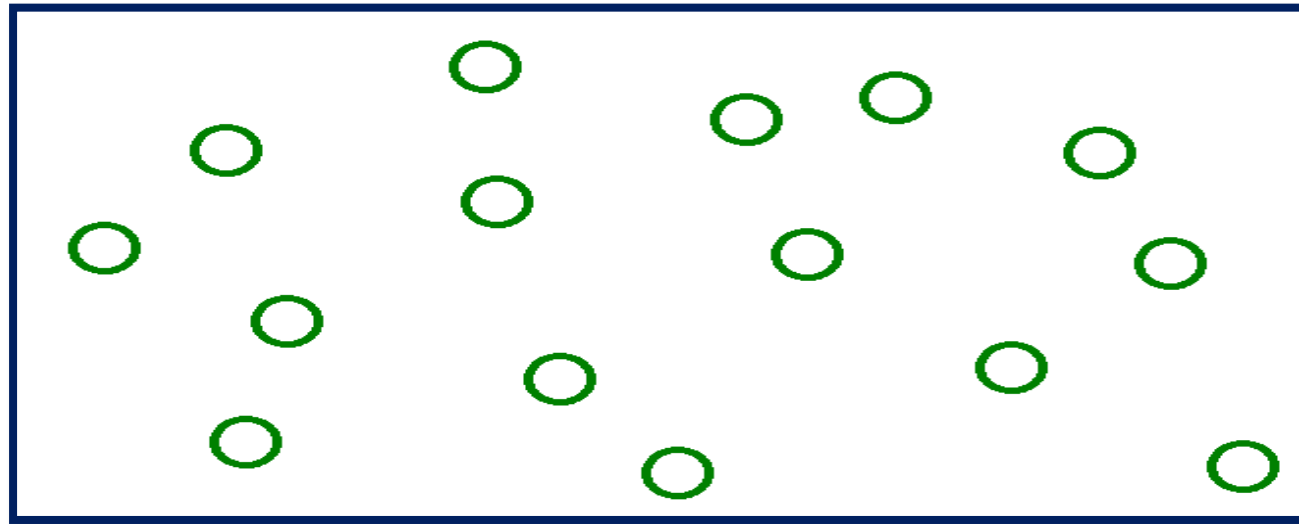
Background

- Research has focused on better understanding the notion of LAL (e.g. Fulcher, 2012) and mapping the LAL of stakeholder groups, including teachers (e.g. Kremmel & Harding, 2015).
- Programmes to improve skills, knowledge, and abilities increasing.
- Shared understanding of the difficulties faced by pre-sessional EAP in-house test developers.

Background

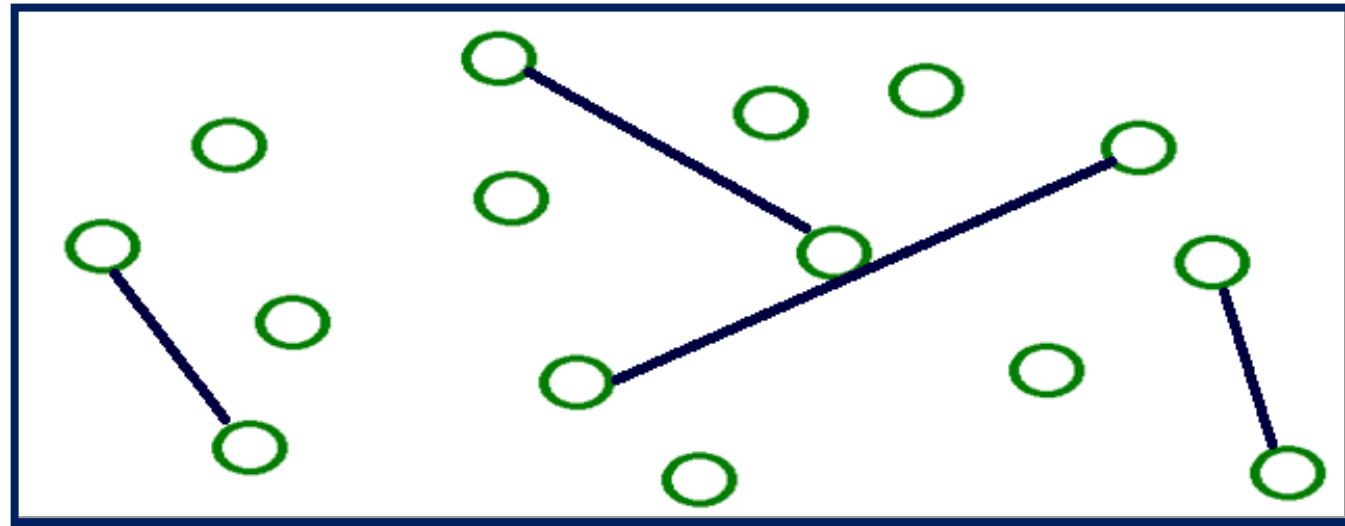
“Testing should be part of every teacher’s skill set, but few teacher training courses cover it adequately. Furthermore, recent (and historical) posts on this list suggest little time is allocated in EAP contexts, and too few staff. And yet the tests we need to produce...are certainly high stakes. So we are in the precarious position of having to produce valid and reliable tests which meet the needs of our students and our other stakeholders, but with restricted resources.” (BALEAP mail list commentator, 2017).

“Most of the assessment work of EAP practitioners is hidden away and thus contributes little to...wider understandings of EAP assessment in practice” (Schmitt & Hamp-Lyons, 2015).



Background

“Most of the assessment work of EAP practitioners is hidden away and thus contributes little to...wider understandings of EAP assessment in practice” (Schmitt & Hamp-Lyons, 2015).



Background

- Defining constructs
- Producing specifications
- Designing items
- Piloting items/tests
- Ensuring equivalence of different test forms
- Deciding grade boundaries
- Training raters
- Ensuring consistent use of assessment matrices

Survey of Assessment Practice

Instrument

- Items based on EALTA Guidelines for Good Practice in Language Testing and Assessment (EALTA, 2006).
- 9 items regarding assessment practices of in-house assessment developers.
- Delivered electronically.

Respondents

- Snow-ball sampling procedure
- 28 EAP institutions responded.

Survey of Assessment Practices

- Are there test specifications?
- Is the test piloted?
(If yes, please give details about both where [e.g. another institution, in-session classes] and the size of the pilot sample compared to test population)
- If there are different versions of the test (e.g. year by year), is the comparability of the different versions measured?
(If yes, then please briefly describe how it is measured)

Survey of Assessment Practices

Provision of Test Specifications

- 22 institutions (79%) reported using test specifications.
- Supports Manning (2013) – lots of examples of good practice.

? “Yes, but they are not written down anywhere.”

Survey of Assessment Practices

Piloting

- 7 institutions (25%) reported piloting assessments.

“Yes, with 4 teachers or so”

“When we can, but it's difficult”

“No. Piloting is considered giving the test”

Survey of Assessment Practices

Equivalence Verification

- 24 institutions reported using multiple tests.
- 19 of these (79%) reported not comparing difficulty of different versions.

“Yes by word of moth amongst colleagues.”

“Not specifically but it's the same people writing so they can try to make sure they are the same level of difficulty.”

“Yes, but only superficially. We need more time to do this.”

Survey of Assessment Practices

Obstacles

- Time, training, resources, and contacts.

“Time, contacts, finances for training”

“It's all about time, which is of course money.”

“Resources - both in terms of not enough staff or time. A lot of assessment "best practices" are kept in mind and we also intend to improve our systems but never seem to get around to it.”

Challenges

- Defining constructs
- Producing specifications
- Designing items
- Piloting items/tests
- Ensuring equivalence of different test forms
- Deciding grade boundaries
- Training raters
- Ensuring consistent use of assessment matrices

Solution

- Defining the construct
- Producing specifications
- Designing
- Piloting
- Ensuring **Collaboration** of different test forms
- Deciding grade boundaries
- Training raters
- Ensuring consistent use of assessment matrices

Example: Setting

- Nottingham Trent University (NTU) pre-sessional with multiple exit points.
- Multifaceted summative assessment. Test of Academic Reading = 20%
- Reading test to date
 - One long text (> 2000 words) & one shorter (< 1500 words)
 - 50-60 questions
 - Not a sustainable design

Example: Design



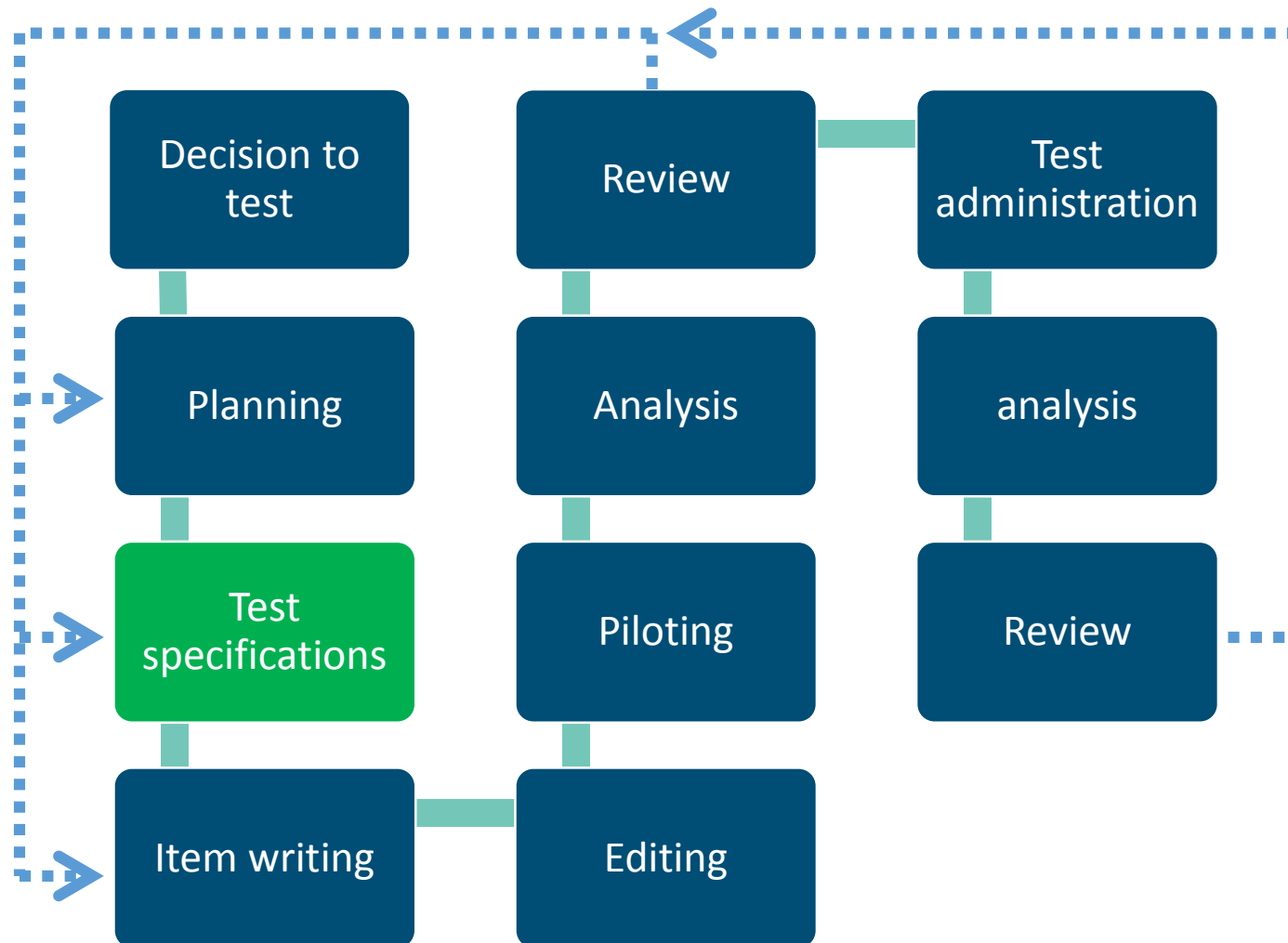
- Tracking data and research (Weir *et al*, 2007) suggested conceptualisation of academic reading was incomplete.
- After **collaboration** with language tester and the literature adopted different model.

Example: Design



- Search for appropriate/applicable theory of reading
- Based on Urquhart and Weir (1998) and Weir and Khalifa (2008).
- Measuring expeditious and careful reading at the global and local level.

Example: Design



- Based on Weir, Huizhong, and Yan (2000)
- Both shorter and longer texts
- Created detailed test specifications (genre, lexical difficulty, illocutionary features, expected reading speed, etc.)

Example: Design



- Sourced lots of texts that matched specifications.
- Wrote items as a team to meet the specifications.
- Piloted them on native English speakers.
- **Collaborated** with another university to pilot all potential items.

Challenges

- Defining constructs
- Producing specifications
- Designing items
- Piloting items/tests
- Ensuring equivalence of different test forms
- Deciding grade boundaries

Challenges

- Defining constructs
- Producing specifications
- Designing items
- Piloting items/tests
- Ensuring equivalence of different test forms
- Deciding grade boundaries

Collaboration: Piloting

- Pilot sample > 70 learners.
 - Took place during mandatory in-session classes.
 - Partner university happy to give students exam practice.
 - Provided feedback on reading level.
- ✓ Has provided us invaluable feedback on test format and new items.
- ? Is this model sustainable? All points of collaboration were facilitated through personal contacts.
- ? If assessment is an achievement test, what will piloting at external institution reveal?

Collaboration: Equivalence

- Test format utilises a variety of interchangeable texts.
 - Item facility of pilot items currently being used as an indicator of equivalence. With anchor items to control for population variance.
-
- ✓ Better than our intuition.
 - ✓ Allows us some confidence in equivalence.
-
- ? More work is needed so we can be confident in our results.

TEXT	LENGTH	TYPE	SOURCE
1	431	BOOK	Skilled Interpersonal Communication: Theory, Research and Practice
2	310	BOOK	Darwin and Facial Expression: A century of research in review
3	220	BOOK	Gesture: How our hands help us think
4	397	BOOK	Social Cognition
5	402	ARTICLE	Charting trends for e-learning in Asian schools
6	148	ARTICLE	Foreign language education and tourism
7	184	ARTICLE	Foreign languages and recruitment in the tourist industry
8	218	ARTICLE	Differences in Uk and non-UK tourism management courses
9	441	ARTICLE	Tourist preferences in holiday selection
10	260	ARTICLE	Low uptake of responsible tourism
11	320	ARTICLE	thrifty gene theory
12	105	ABSTRACT	Laptop multitasking hinders classroom learning for both users and nearby peers
13	134	ABSTRACT	The Determinants of International Mobility of Students
14	118	ABSTRACT	Journalism Students' Professional Views in Eight Countries: The Role of Motivations, Education, and Gender
15	107	ABSTRACT	Reconsidering Brainstorming
16	95	ABSTRACT	International students in English-speaking universities: Adjustment Factors
17	139	ABSTRACT	Exploring the Impact of Study Abroad on Students' Intercultural Communication Skills: Adaptability and Sensitivity
18	136	ABSTRACT	Defining and measuring employability
19	128	ABSTRACT	Measuring the Gender Gap on the Internet
20	529	BOOK INTRO	Understanding and developing student engagement
21	184	BOOK	Language
22	462	TEXTBOOK	The Aral Sea
23	341	BOOK	Investigating English Style
24	394	BOOK	Learning to talk - Varieties of English
25	376	BOOK	Sociology: A Global Introduction
26	117	BOOK	Research Methods in Education

	TEXT	COMPONENT	#	Question type	Question focus
Ethical tourism	9	ELSK	1	Multiple choice	Topic
Ethical tourism	9	ELSE	1	Multiple choice	Sentence selection
Ethical tourism	9	ELSC	1	Multiple choice	Data comprehension
Ethical tourism	9	ELSC	2	Multiple choice	Data comprehension
Ethical tourism	9	CLS	1	Short Answer	Pronominal reference
Ethical tourism	10	ELSK	1	Multiple choice	Summary
Ethical tourism	10	ELSK	2	Multiple choice	Propositional structure
Ethical tourism	10	ELSE	1	Multiple choice	Sentence selection
Ethical tourism	10	ELSC	1	Multiple choice	Word identification
Ethical tourism	10	CLT	1	Sentence Input	
Ethical tourism	10	CLL	1	Multiple choice	Synonym
Obesity	11	ELSK	1	Multiple choice	Summary
Obesity	11	ELSE	1	Multiple choice	Sentence selection
Obesity	11	ELSE	2	Multiple choice	Sentence selection
Obesity	11	ELSE	3	Multiple choice	Sources
Obesity	11	CLL	1	Multiple choice	Synonym
IT & ED	12	ELSK	1	Multiple choice	Title
International students	13	ELSK	1	Multiple choice	Title
Journalism education	14	ELSK	1	Multiple choice	Title
Brainstorming	15	ELSK	1	Multiple choice	Title
International students	16	ELSK	1	Multiple choice	Essay question
Study abroad	17	ELSK	1	Multiple choice	Essay question
Employability	18	ELSK	1	Multiple choice	Essay question
Gender and Internet	19	ELSK	1	Multiple choice	Essay question
Student engagement	20	ELSE	1	Multiple choice	Paragraph match
Student engagement	20	ELSE	2	Multiple choice	Paragraph match
Translation	21	ELSK	1	Multiple choice	Author's voice

Collaboration: Equivalence

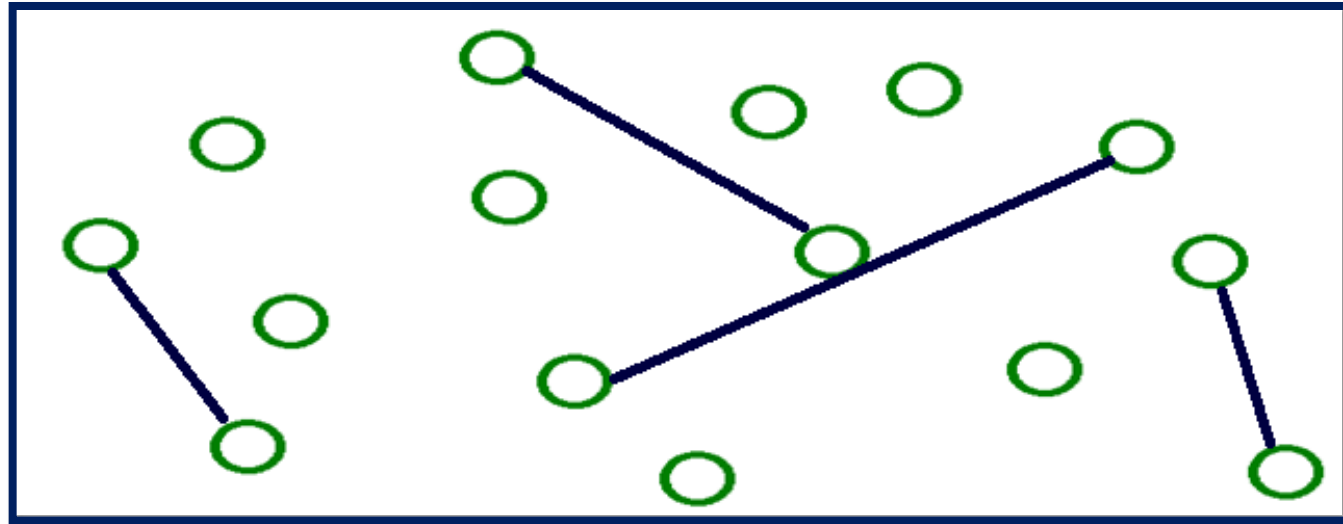
- Test format utilises a variety of interchangeable texts.
 - Item facility of pilot items currently being used as an indicator of equivalence. With anchor items to control for variance between pilot populations.
-
- ✓ Better than intuition.
 - ✓ Allows us some confidence in equivalence.
-
- ? More work is needed so we can be confident in our results.

Collaboration: Cut-scores

- In collaboration with language tester, several methods are being considered.
- Goal is to set meaningful cut-scores based on data.

Reflections

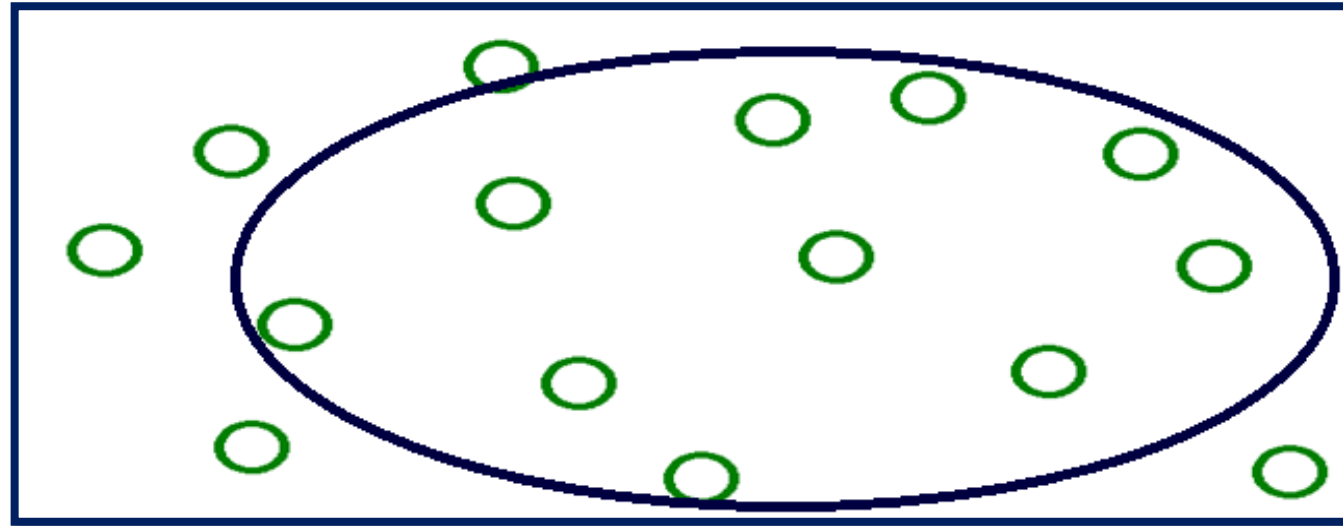
- Much of this design facilitated by collaboration, but the current model is not sustainable.



Reflections

- Need a more sustainable model.

? Informal network of institutions willing to pilot and give feedback on instruments.



Reflections

- This is not a new idea (I told you!):

“Working through teacher associations or cross-university partnerships, groups of individuals from different universities could join together to initiate collaboration with language testing specialists and thereby develop the critical mass of testing expertise required to create and maintain high quality assessments” (Schmitt & Hamp-Lyons, 2015) .

“It all goes back to that community and sort of the collaboration thing”
(Manning, 2016).

Circle all relevant answers.

(1 mark)

1. Is such a network _____?

- a) feasible
- b) useful
- c) sustainable
- d) desirable

samuel.barclay@ntu.ac.uk

