

# Construct Validity & Specifications in Language Assessment

**EAP in the North September 2018**

Dr Anthony Manning



# Overview

## Session Structure:

### 1. Scene setting

- The people affected
- Defining language assessment
- Your context
- Ethics of assessment
- Assessment literacy

### 2. Introducing Construct validity

- Weakness in construct validity
- ILOs and assessment

### 3. Using test specifications

### 4. Test purpose and function

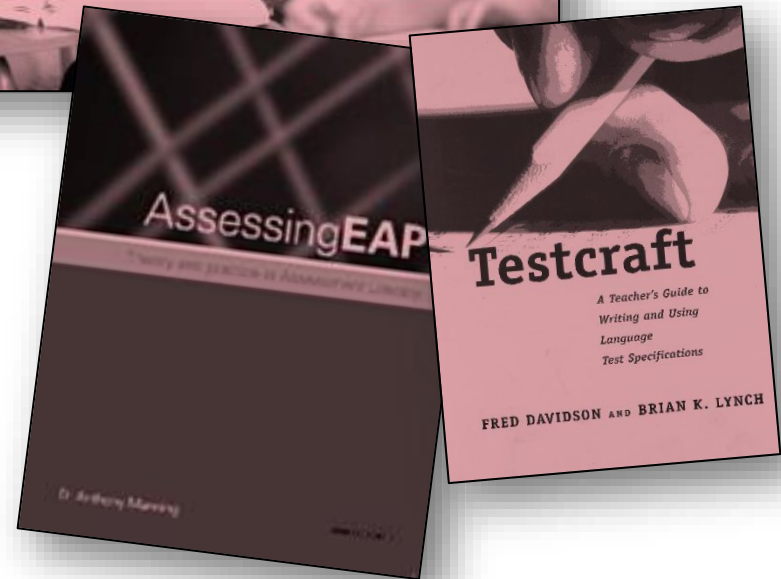
### 5. Sampling

### 6. Piloting and trialling

### 7. Good practice tips

## ● Key Reference Material:

- Testcraft
- Assessing EAP: theory and Practice in Assessment Literacy
- Other references from the field



# *The start of a new EAP course...*

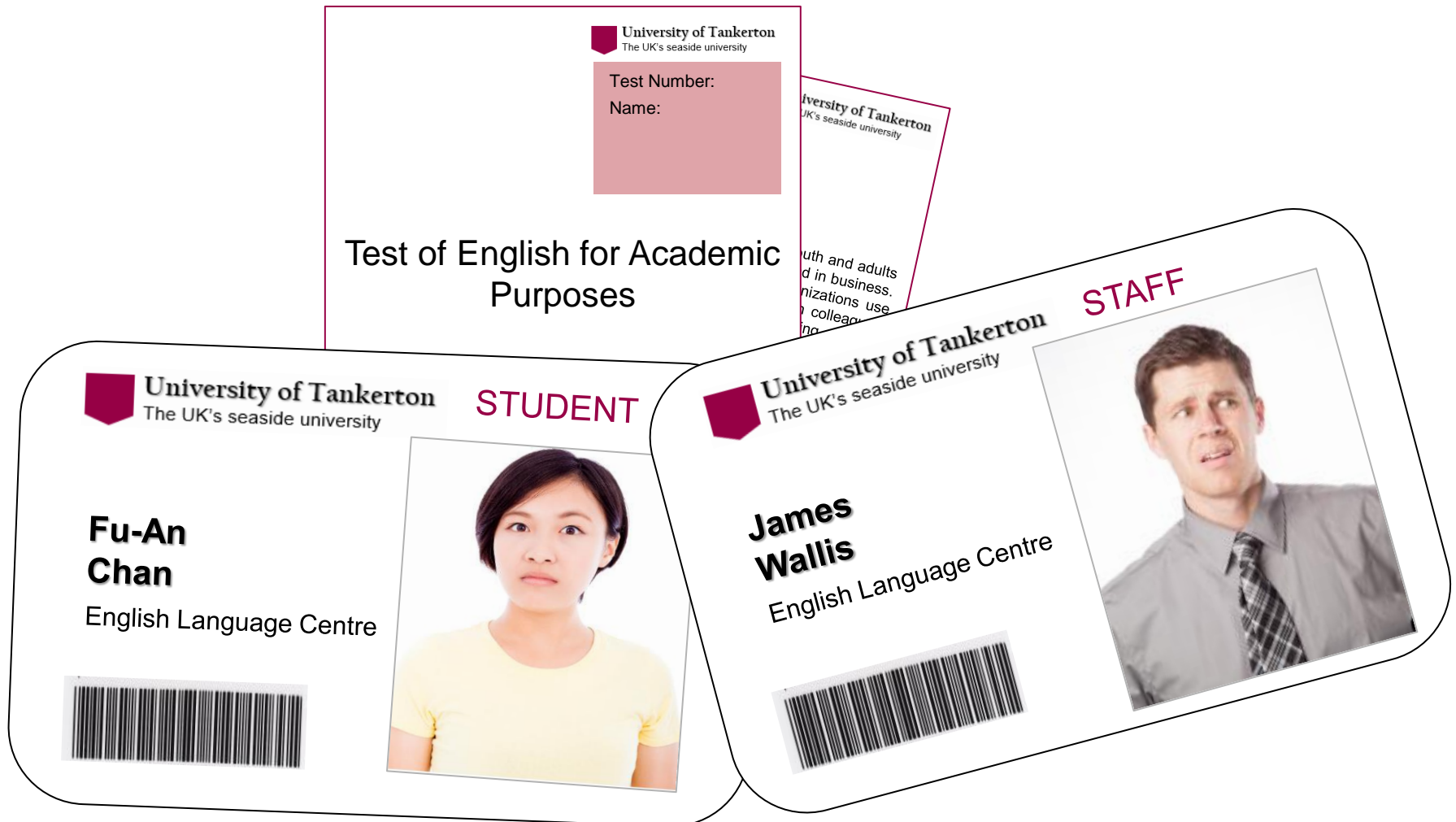


# *Adjusting to EAP in context...*

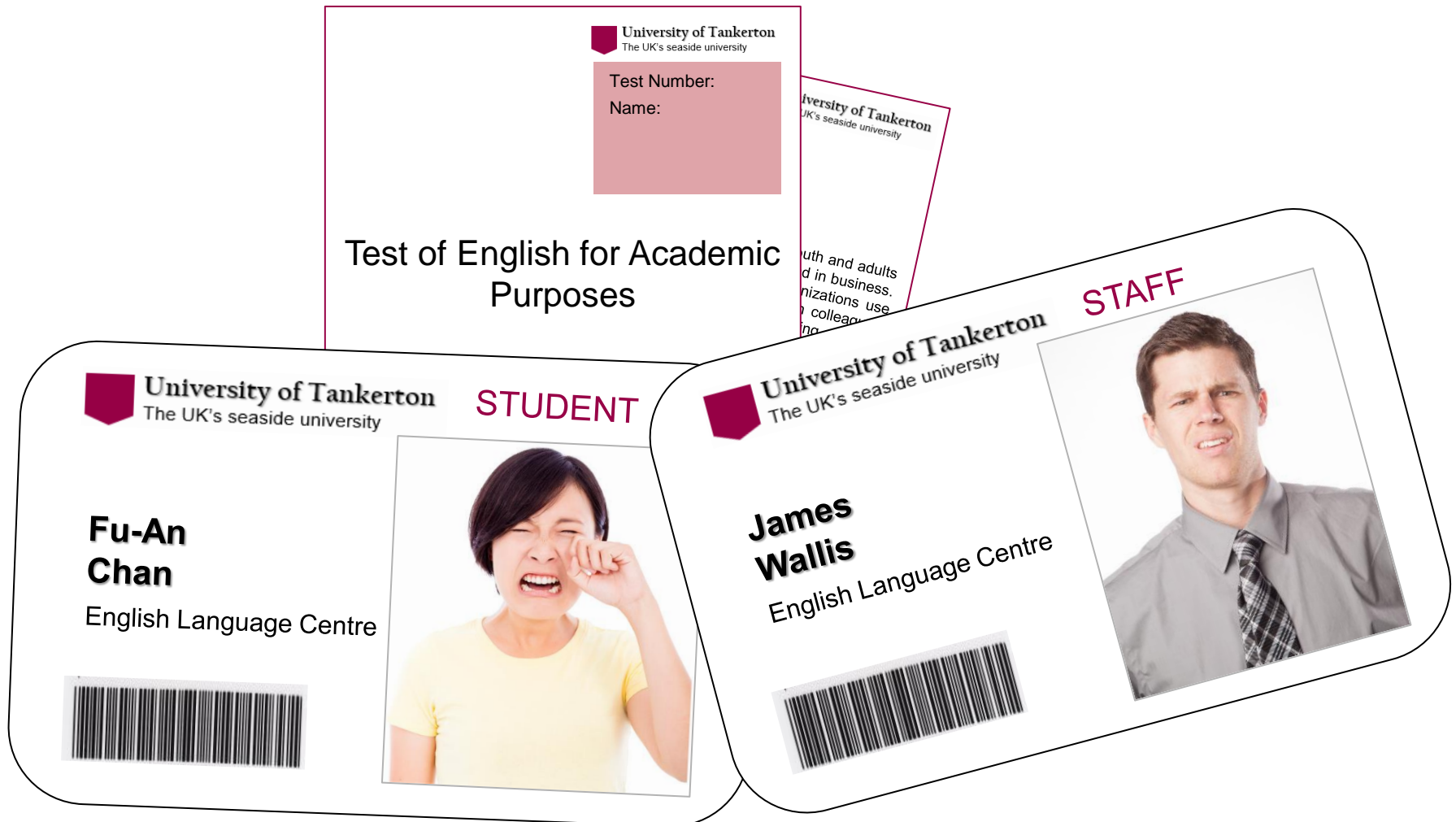




# Preparing for the first assessments...



# Taking account of the results...



## Ongoing impact...



# 1. Setting the scene – Defining language assessment

- In the context of language teaching and learning, 'assessment' refers to the **act of collecting information and making judgments about a language learner's knowledge of a language and ability to use it.** (Chappelle and Brindley, 2010)
- **Critical Language Testing**

The work of Shohamy (1998, 2001) and McNamara & Roever (2006) warn of the power of tests and the potential harm which can be incurred if tests and their results are misused or misinterpreted.





# 1. Setting the scene - your assessment context

## This section will:

- Set the scene by discussing the different contexts in which you test or assess EAP
- Start to consider the challenges which you and your students face in EAP assessment



Why are my students studying EAP and why do they need to be assessed?

Which skills do my students need to develop and demonstrate for use in their future use of EAP?

How can I continue to enhance the quality of my EAP assessments?

Teaching and assessing EGAP or ESAP? How can I reflect this meaningfully in the syllabus and assessment?

Which other challenges do I face in teaching and assessing EAP?

What are the consequences of poor quality EAP assessment?

What type of assessment tools and sources should I use?



# 1. Setting the Scene- Ethics of assessment

## This section will:

- introduce you to a range of ethical considerations associated with language assessment
  - invite you to explore the ethicality of your own practice and experience
- 
- Language tests should be given a health warning (Spolsky 1981, p. 20).
  - Account needs to be taken on the gate-keeping function of certain language-related tests: institutionally, nationally and internationally (Flowerdew & Peacock, 2001, p. 192).
  - Shohamy (1998, p. 332, 2001) advocates a critical approach to language testing which acknowledges that the act of testing is not neutral.
  - Bachman (1990, p. 279) notes that 'tests are ...virtually always intended to serve the needs of an educational system or of society at large.'
  - McNamara & Roever (2006, p. 8) note that testers need to engage in debate on the consequential application of their tests.



# 1. Setting the scene -Assessment literacy

## This section will:

- demonstrate the importance of assisting stakeholders to understand what the results of language tests show
  - highlight the importance of language Assessment Literacy amongst different stakeholder groups
- 
- The term Assessment Literacy in language assessment recognises that there is a need to describe what language teachers need to know about assessment matters (Inbar-Lourie, 2008; Malone, 2011; Stiggins, 1991; Taylor, 2009)
  - This need for a dynamic response resonates with Popham's (2001, 2006) call for proactivity in the training of educators and other stakeholders.
  - There is a need for assessment literacy across stakeholder groups whether people are involved in the process of selecting, administering, interpreting and sharing results of large-scale tests produced by testing or examination boards, or in producing, marking, analyzing and enhancing in-house or classroom-based assessments (Taylor, 2009, p. 24).



## 2. Introducing construct validity

### This section will:

- introduce you to the concept of construct validity
  - explain how construct validity can assist in improving tests and assessments in EAP
- 
- A construct is an area of ability or skill in language use. Although this term may seem quite simple, however, when we comes to actually construct it it is actually quite difficult to achieve.
  - Wigdor and Garner's definition of construct validity (1982, p. 62) describes construct validity as '**... a scientific dialogue about the degree to which an inference that a test measures an underlying trait or hypothesized construct is supported by logical analysis and empirical evidence**'.
  - Messick (1989) describes construct validity as a unitary concept with various features. This view prioritizes concern for creating tests which include accurate reflections of relevant constructs, it also focuses on usage of test results, given the consequences for people's lives.





## 2. Weakness in construct validity

- Consider the construct validity of the following set of test items:

### Question 5

Read the three sentences below and rewrite them using the passive voice, so that they appear more academic in style.

- 5a) I undertook this experiment in order that I could determine whether the level of pollution had changed in the stream since I collected my first sample two weeks earlier. 1 mark
- 5b) I chose a mixed-methods approach to my research as I felt that I needed to use both quantitative and qualitative data in my investigations. 2 marks
- 5c) I interviewed 25 people so that I could obtain a more representative view of professionals in my field. 1 mark

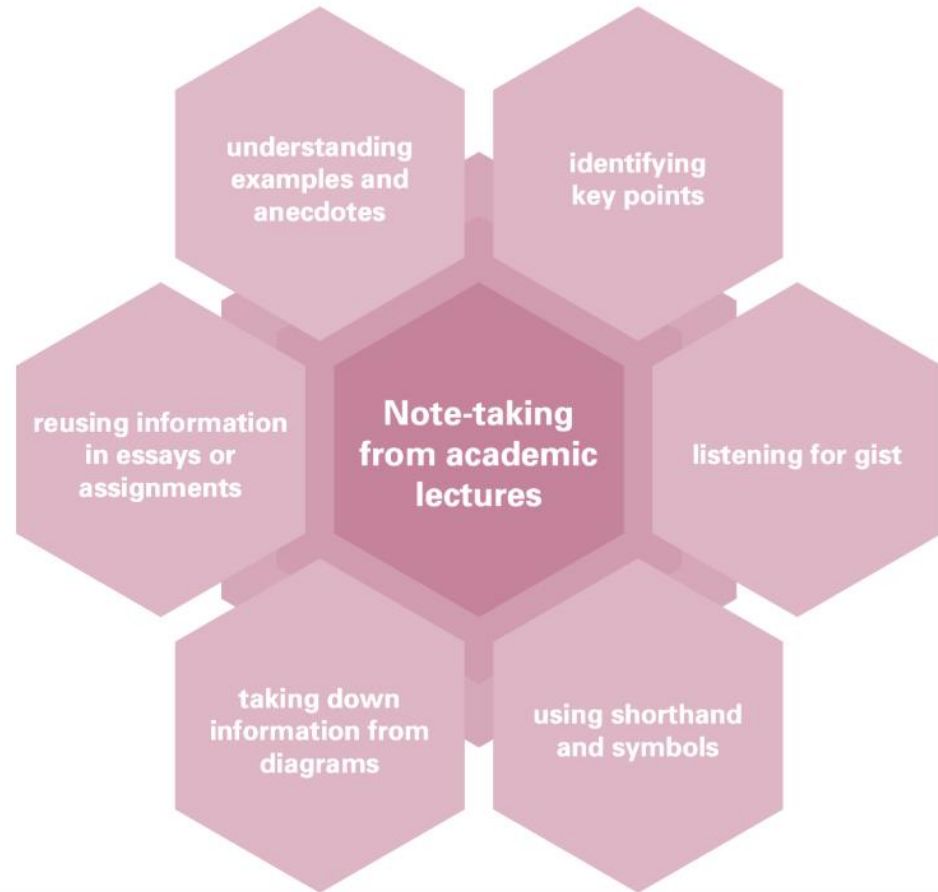


- How well do you think the construct of 'passive voice' in academic writing' has been realized in the set of items above?
- What are the possible weaknesses in how this set of items has represented the construct in focus?
- What does it really show if a student gets a high or a low score after completing this set of items?
- How could this set of items be improved?

## 2. Construct validity – ILOs and Assessment

Look at the different dimensions of the construct of note taking

- How would you represent the construct of Academic Writing? Which skills or constructs would you include?
- How would you build the different aspects of Academic Writing into your syllabus and your test?
- How do you ensure that your assessments assess your domain appropriately?



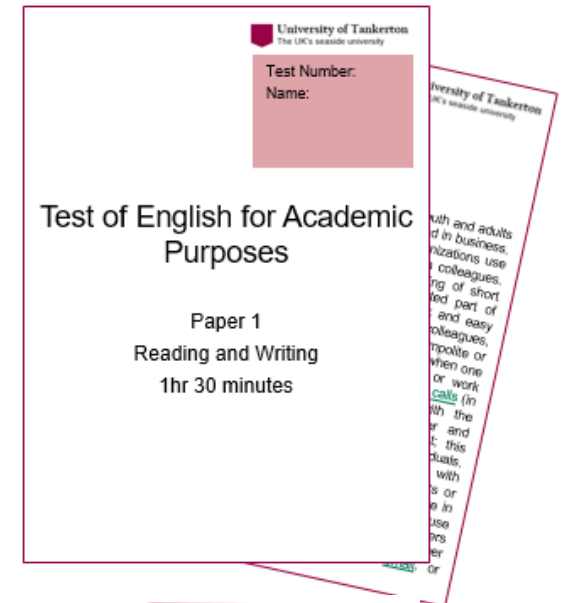
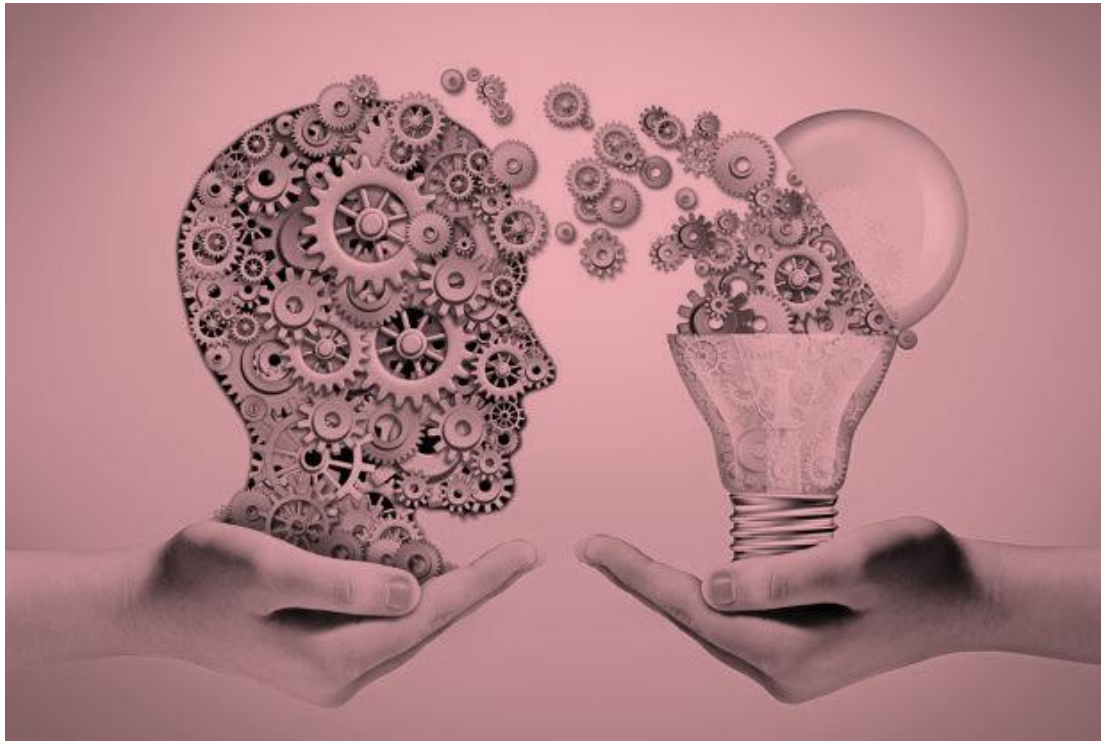
### 3 Using Test Specifications

- In language testing, the term specification is used to describe a generative blueprint document through which alternative versions of an assessment or assessment task can be created (Davidson & Lynch, 2002, p. 4).



# 3. Using Test Specifications – reverse engineering

- When working from pre-existing assessments or tests, critical reverse engineering is advocated in order to avoid divergence from the main aims of the assessment and the constructs which are considered essential to measure.





### 3. Using Test Specifications

- The structure below is an example of how part of a specification could be structured for a writing test item. Do you already use a test specification when you write EAP tests? If so how does it help? If not why not? What do you think of this example framework?



<b>Specification Number</b>	
<b>Paper</b>	
<b>Section</b>	
<b>Item Number</b>	
<b>General Description</b> (description of the behaviour being assessed)	
<b>Prompt Attributes</b> (detailed description of what the the test taker will see in the test tem)	
<b>Response Attributes</b> (detailed description of how the student should provide the answer linked to criteria)	
<b>Example Item</b> (Example item or task reflecting the spec)	
<b>Source/ sampling Information</b>	

### 3. Using Test Specifications

- The structure below is an example of how part of a specification could be structured for a writing test item. Do you already use a test specification when you write EAP tests? If so how does it help? If not why not? What do you think of this example framework?

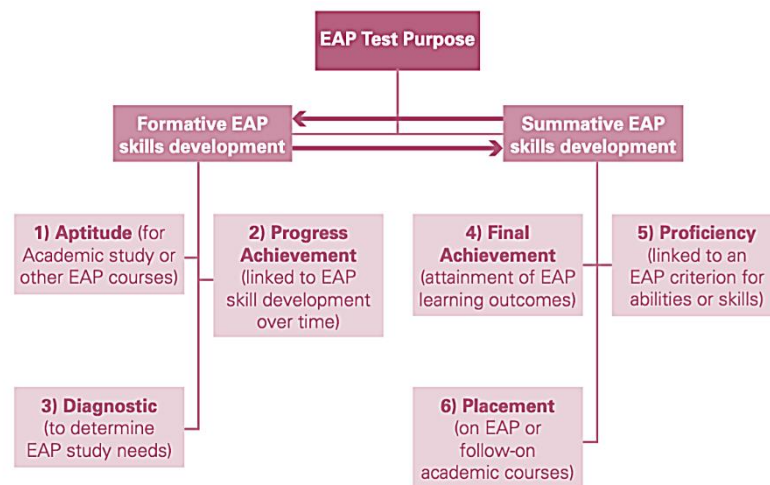


<b>Specification Number</b>	25 (Academic English Level 1)
<b>Paper</b>	2 (Writing)
<b>Section</b>	2
<b>Item Number</b>	1
<b>General Description</b> (description of the behaviour being assessed)	The ability to write a discursive essay which demonstrates critical thinking, relevant to .
<b>Prompt Attributes</b> (detailed description of what the the test taker will see in the test tem)	A short 300-word essay, to be written over a period of 30-45 minutes
<b>Response Attributes</b> (detailed description of how the student should provide the answer linked to criteria)	Write an essay with an introduction, main body and conclusion which presents a balanced argument on the topic taking into consideration different perspectives which are relevant.
<b>Example Item</b> (Example item or task reflecting the spec)	Write an essay of 300 words for a university lecturer on the following topic: <b>to what extent do you agree that University education should be accessible to all members of society</b>
<b>Source/ sampling Information</b>	Essay title negotiated with UG module convenors from Humanities and social Sciences – suitable topic of cross-over interest

## 4. Test purpose and function

### This section will:

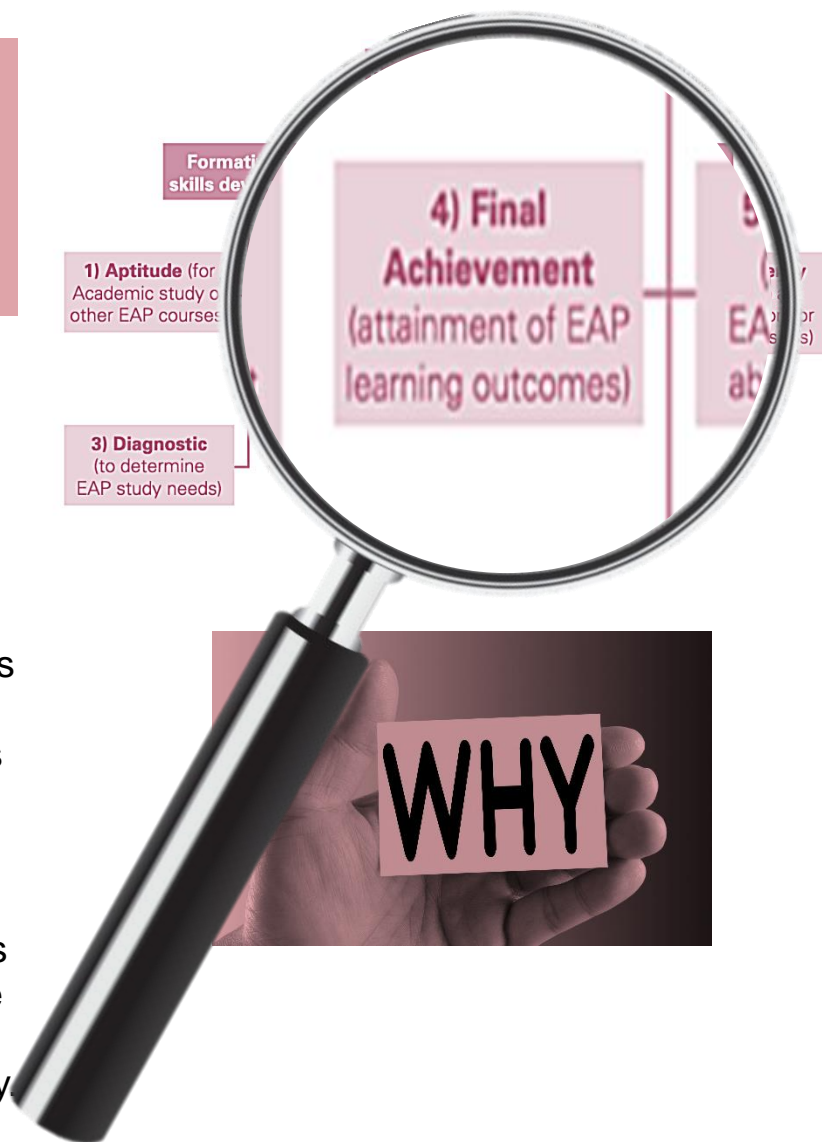
- Relate the use of specifications and construct validity to a range of purposes and functions of assessment in EAP summative assessment in EAP
- As clarified by Berry (2008, p. 13), assessment is used for a range of different reasons, which can be categorized in a series of differing ways. In order for language assessment to be appropriate, suitable assessment tools need to be selected in line with the particular assessment objectives.
- Alexander et al. (2008, p. 307) describe the complexities of trying to separate the functions of tests into formative and summative categories. Some summative EAP tests may also have a formative function, whilst a test designed for EAP proficiency measurement may also yield diagnostic information.
- By identifying learning outcomes, or stated expectations of what someone will have learned and how they will be assessed, student centred learning can be promoted along with the improvement of curriculum and pedagogy. (Driscoll and Wood 2007, p.5)



## 4. Test purpose and function

### This section will:

- Relate the use of specifications and construct validity to a range of purposes and functions of assessment in EAP summative assessment in EAP
- As clarified by Berry (2008, p. 13), assessment is used for a range of different reasons, which can be categorized in a series of differing ways. In order for language assessment to be appropriate, suitable assessment tools need to be selected in line with the particular assessment objectives.
- Alexander et al. (2008, p. 307) describe the complexities of trying to separate the functions of tests into formative and summative categories. Some summative EAP tests may also have a formative function, whilst a test designed for EAP proficiency measurement may also yield diagnostic information.
- By identifying learning outcomes, or stated expectations of what someone will have learned and how they will be assessed, student centred learning can be promoted along with the improvement of curriculum and pedagogy (Driscoll and Wood 2007, p.5)





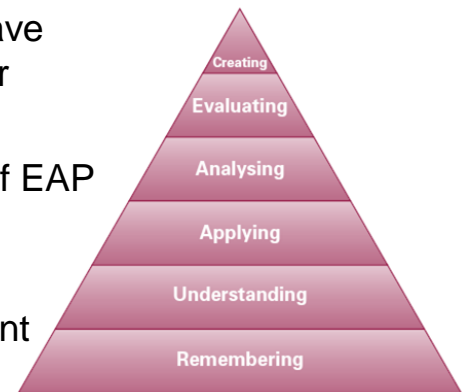
# 5. Sampling what needs to be assessed

## This section will:

- Introduce you to the challenges involved in sampling domains for the purpose of language assessment
- Encourage the implementation of certain strategies to assist with the sampling process



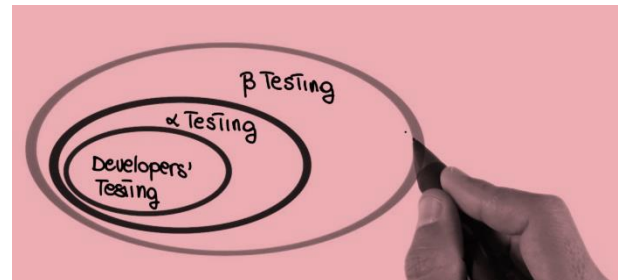
- Kane (2012, pp. 41–42) explains, identifying the range of performances to include in a target subject area, and deciding on the emphasis can be complicated.
- Approaches to needs analysis have been developed (e.g. Jordan, 1997) have relevance to the sampling process, by helping to identify which key skills for assessment can be identified.
- Weir (1983) believed that content validity could be achieved in the testing of EAP through analyzing test takers and their needs, and using this information to determine the content of tests.
- However, Fulcher (1999 p. 221) argues that concerns for authenticity content should not remove focus from how valid inferences from tests are drawn
- Driscoll and Wood (2007, p. 5) support identifying learning outcomes, or stated expectations of what someone will have learnt and how they will be assessed,
- Green (2007, p. 48) explains that Blooms taxonomy can be harnessed, in academic writing, to both build and assess student understanding- also used to inform criteria.



## 6. Piloting and trialling

### This section will:

- Provide mechanisms for enhancing tests before they are used
- Identify approaches to trialling and piloting EAP tests and assessments



- Fulcher and Davidson (2007, p. 81) compare the trialling and piloting of language tests to the manufacturing industry and refer to two stages of rapid prototyping, which involves alpha testing and beta testing. (Davies et al., 1999, p.150)
  - Alpha- proofreading, peer review, external examiners, subject specialists ...
  - Beta- results analysis, test-taker pre-test comments, post-test comments, test annotations...
- Skills in pre-testing relate directly to the area of assessment literacy associated with avoiding and preventing test-related problems
- Bachman and Palmer (1996,p.234) suggest three phases of pre-testing, as indicated in Table 1
- In High-stakes situations, pre-testing is very important as it is part of the process of supporting validity.

Table 1: Steps in pre-testing (Bachman & Palmer, 1996, p. 244)

Step 1	An informal stage where individuals (experts or representative students) and small groups provide qualitative feedback on a test regarding potential problems, specific tasks and instructions.
Step 2	The collection of more quantitative feedback on the test from larger groups of representative students or expert colleagues.
Step 3	A field trial conducted under operational conditions using the same administration procedures as when the test will be given for its intended measurement purposes. This field test stage, however, is only used to assess test usefulness and not to make formal inferences about test takers.

## 7. Good practice tips

- Define your test purpose and use that as your starting position
- Build your questions/items so that they reflect the constructs that you really want to measure and that are important for students' and stakeholders' future usage
- Create or build a test specification and update it when necessary
- Be wary of lone measures and other factors affecting reliability.
- Keep track of your items through good record keeping and item banking
- If you have to reverse engineer, then do so with great care and a critical eye
- Remember that good assessments are built through collaboration and consensus
- Test your test, with colleagues and if possible representative test-takers
- Always look for improvements to your model and actively seek and act on feedback
- Correct any errors as soon as you spot them, or they are likely to linger
- Be sure that your marking key and criteria are sufficiently strong and universally interpreted
- Train your markers, through standardization calibration activities and undertake group moderation activities
- Ask your test takers and stakeholders for their views and suggestions for improvement
- Make user-friendly tests- rubrics, proofreading and formatting
- Take your time and lobby for more if it isn't immediately available
- Every little helps!



# References A

- Alexander, O., Argent, S., & Spencer, J. (2008). *EAP essentials: A teacher's guide to principles and practice*. Reading: Garnet.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Berry, R. (2008). *Assessment for learning*. Hong Kong London: Hong Kong University Press; Eurospan
- Biggs, J. B., & Tang, C. S. (2011). *Teaching for quality learning at university: what the student does* (4th ed.). Maidenhead: McGraw-Hill/Society for Research into Higher Education/Open University Press.
- Chappelle, C.A. & Brindley, G. (2010) Assessment. In Schmitt, N (ed). An introduction to Applied Linguistics. Abingdon, Oxon.: Hodder and Stoughton
- Davidson, F., & Lynch, B. K. (2002). *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven; London: Yale University Press.
- Davies, A., Brown A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999) Dictionary of Language Testing. Cambridge: Cambridge University Press.
- Driscoll, A., & Wood, S. (2007). *Developing outcomes-based assessment for learner-centered education: A faculty introduction*. Sterling, Virginia: Stylus Publishing.

# References B

- Flowerdew, J., & Peacock, M. (2001). *Research perspectives on English for academic purposes*. Cambridge: Cambridge University Press.
- Fulcher, G. (1999). Assessment in English for academic purposes: Putting content validity in its place, *Applied Linguistics*, 20(2), 221–236.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London: Routledge.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press.
- Hughes, A. (2003). Testing English for University Study. *ELT Documents: 127*. Oxford: Modern English Press
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402.
- Jordan, R. R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge: Cambridge University Press.
- Kane, M. (2012). Articulating a validity argument. In F. Davidson & G. Fulcher (Eds.), *The Routledge Handbook of Language Testing* (pp. 34–47). Oxford: Routledge.
- Knight, P. (1995). *Assessment for learning in higher education*. London: Kogan Page.
- Manning, A. (2016). *Assessing EAP: Theory and Practice in Assessment Literacy*. Reading: Garnet



# References C

- Malone, M. E. (2011). *Assessment Literacy for Language Educators* (CAL Digest, October 2011). Washington, DC: Center for Applied Linguistics.
- McNamara, T. & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed.). New York: American Council on Education.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: ASCD.
- Popham, W. J. (2006). Needed: A dose of assessment literacy. *Educational Leadership*, 63(6), 84–85. Retrieved from [http://www.ascd.org/publications/educational\\_leadership/mar06/vol63/num06/Needed@-A-Dose-of-Assessment-Literacy.aspx](http://www.ascd.org/publications/educational_leadership/mar06/vol63/num06/Needed@-A-Dose-of-Assessment-Literacy.aspx)
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow: Longman.
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24(4), 331–345.
- Stiggins, R. J. (1991). Assessment Literacy. *The Phi Delta Kappan*, 72(7), 534.
- Spolsky, B. (1981). Some ethical questions about language testing. In C. Klein-Braley & D. K. Stevenson (Eds.), *Practice and problems in language testing I*, (pp. 5–21). Frankfurt: Peter Lang.
- Taylor, L. (2009). Developing Assessment Literacy. *Annual review of Applied Linguistics*, 29, 21–36.

# References D

- Weir, C. J. (1983). *Identifying the language problems of overseas students in tertiary education in the United Kingdom* (PhD Thesis, University of London Institute of Education).
- Wigdor, A. K., & Garner, W. R. (1982). *Ability testing: Uses, consequences and controversies*. Washington, D.C.: National Academy Press.

# THE UK'S EUROPEAN UNIVERSITY



[www.kent.ac.uk](http://www.kent.ac.uk)

University of  
**Kent**