

## Authenticity in assessment: What does it mean?

1 April 2019 – IATEFL TEASIG pre-conference event  
Dr John Pill

### An authenticity 'gap'?

*I had to look at a diagram of the process and describe how to make it – there was a time limit, and it was very difficult to achieve in almost perfect English – how to make jam.*

#### IELTS Academic Writing Task 1

*describe and explain data, describe the stages of a process, how something works or describe an object or event*

(IELTS information for candidates, 2019)

<https://www.bbc.co.uk/news/health-41010021> 22 August 2017  
<https://www.ielts.org/about-the-test/publications>



**BBC NEWS**  
Home UK World Business Politics Tech Science Health Family & Education

**Health**

### Do I have to understand jam-making to be a nurse?

By Hugh Pym  
Health editor

© 22 August 2017

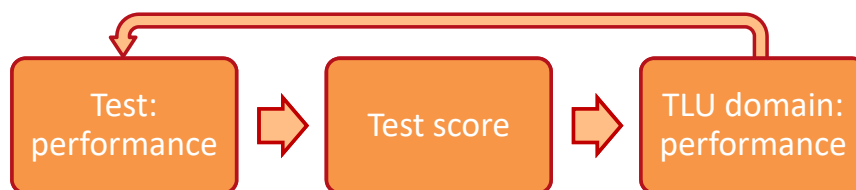
Polina Ralutin thinks the language test is too difficult

Polina Ralutin was an experienced nurse in the Philippines and keen to take up a post in the NHS, at the Lister hospital, in Stevenage.

## Basic assumptions

*specificity encompasses both authenticity and purpose*  
(Knoch & Macqueen, 2016)

real world = *target language use (TLU) domain*  
(Bachman & Palmer, 1996)



Authenticity promotes test-taker motivation and positive washback

## Authenticity of topic and text

Authentic for:

*Test takers*

students starting university

*Context of test*

introductory course in finance,  
economics

*Purpose of test*

readiness to study in English, to read  
and interpret relevant texts

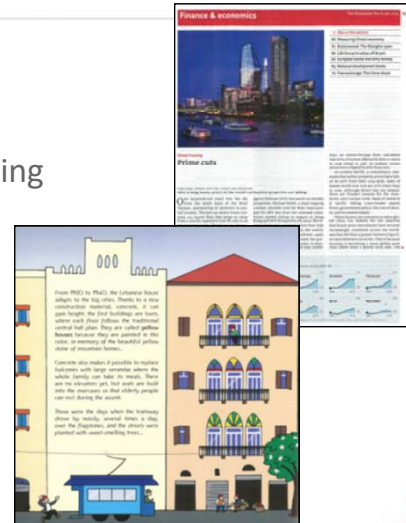


The Economist. 9 March 2019. p. 65

## Authentic texts in tests

### Challenges to authenticity

- Text changes when it is re-purposing as a test
- Change of context may require new wording, editing (assumptions of shared knowledge are lost)
- Text may not be suitable for test use
- Text formatting or presentation may change
- Text can be found on the Web quickly
- Copyright law affects text use



## Authenticity of task

### Skills

The tasks may require the test taker to:

- gain a sense of the text as a whole (gist)
- locate and distinguish main ideas and important information
- locate specific details
- summarize or paraphrase information
- recognize the use of cohesive and other structural devices in written text
- locate sequences, contrasts, causes and effects in the text
- make inferences/draw a conclusion based on information provided
- distinguish between fact and opinion
- infer the meaning of unfamiliar lexical items from their context
- recognize the writer's intention in context (e.g., sarcasm, advising)
- identify audience, purpose and genre of a text



## Authenticity of task

In paragraph III, the writer implies

- (A) encourage innovation.  
 (B) are often overwhelming  
 (C)\* demand many decisions  
 (D) simplify content-sharing

0	_____ E _____
1	_____
2	_____
3	_____
4	_____
5	_____

- A ... has less  
 B ... gives you  
 C ... teaches  
 D ... offers to  
 E ... improves  
 F ... helps you  
 G ... focuses

1. When did bronze vessels first appear?

2. Who were the vessels offered to?

3. Where were many of the vessels found?



A superficially inauthentic task  
 may prompt authentic  
 cognitive processes  
 (interactional authenticity)

1. _____	➔	2. _____
	➔	3. _____

## Authenticity of task

*Graded Examinations in Spoken English (GESE)*  
 In the interactive phase:

- Take responsibility for the direction and maintenance of the interaction
- Utilise turn-taking conventions to ensure that the interaction flows and develops naturally
- Relate his or her own contributions closely to those of the examiner

### Grade 10

The examiner will introduce the Interactive phase by saying:

In this task, I'll start by telling you something. You'll have to ask me questions to find out more information and make comments. It's your responsibility to maintain the conversation. Are you ready?

### Language functions

- ▶ Developing an argument
- ▶ Defending a point of view
- ▶ Expressing beliefs
- ▶ Expressing opinions tentatively
- ▶ Summarising information, ideas and arguments
- ▶ Deducing

**TRINITY**  
 COLLEGE LONDON

### Grade 10 Interactive prompts

- At my niece's school, homework is optional. Students only have to do it if they want to. I've even heard there's a move to get rid of it altogether.
- Many people think that we should spend more time enjoying ourselves and less time trying to make money, but I'm not sure you can do one without the other.
- I was listening to a talk about architecture the other day and the view presented was that all modern architecture is ugly.
- A lot of people complain about graffiti in cities these days, but I'm not sure if I agree with them.
- Recently there seems to have been a reaction against our technologically advanced society, with some people preferring to return to a simpler lifestyle. I think there's something to be said for this.
- People often say that academic qualifications are the key to a child's future success. I'm not sure I completely agree.

<https://www.trinitycollege.com/resource/?id=5755> <https://www.trinitycollege.com/resource/?id=5731>

### OCCUPATIONAL ENGLISH TEST

**WRITING SUB-TEST: NURSING**  
**TIME ALLOWED: READING TIME: 5 MINUTES**  
**WRITING TIME: 40 MINUTES**

Read the case notes and complete the writing task which follows.

**Notes:**

Mr Gerald Baker is a 70-year-old patient on the ward of a hospital in which you are Charge Nurse.

**Patient Details:**

Marital Status: Widower (8 years)  
Admission Date: 3 September 2010 (City Hospital)  
Discharge Date: 7 September 2010

**Diagnosis:**

Left Total Hip Replacement (THR)  
Ongoing high blood pressure

**Social Background:**

Lives at Greywatts Nursing Home (SNH) (8 years)  
No children  
Employed as a radio engineer until retirement aged 65  
Now aged-pensioner  
Hobbies: chess, ham radio operator  
Sister, Dawn Mason (86), visits regularly; v supportive  
- plays chess with Mr Baker on her visits  
No signs of dementia obtained

**Medical Background:**

2008 - Osteoarthritis requiring total hip replacement surgery  
1999 - Hypertension (ongoing management)  
1995 - Colles fracture, ORIF

TURN OVER

**Medications:**

Aspirin 100mg twice (recommended post-operatively)  
Rampri 6mg twice  
Paracetamol 500mg (paracetamol) 2 qid pm

**Nursing Management and Progress:**

daily dressings surgical incision site  
Range of motion, stretching and strengthening exercises  
Occupational therapy  
Staples to be removed in two wks (21/9)  
Also, follow-up FBE and UEC tests at City Hospital Clinic

**Assessment:**

Good mobility post-operation  
Weight-bearing with use of wheeled-walker, walks length of ward without difficulty  
Post-operative disturbance in time and place during recovery, possibly relating to anaesthetic - continued observation recommended  
Dropped Hb post-operatively (to 72) requiring transfusion of 3 units packed red blood cells; Hb stable (112) on discharge - ongoing monitoring required for anaemia

**Discharge Plan:**

Monitor medications (Paracetamol)  
Preserve skin integrity  
Continue exercise program  
Equipment required: wheeled-walker, wedge pillow, toilet raiser. Hospital to provide walker and pillow. Hospital social worker organised 2-wk hire of raiser from local medical supplier.

**Writing task:**

Using the information in the case notes, write a letter to Ms Samantha Bruin, Senior Nurse at Greywatts Nursing Home, 27 Station Road, Greywatts, who will be responsible for Mr Baker's continued care at the Nursing Home.

**In your answer:**

- expand the relevant notes into complete sentences
- do not use note form
- use letter format

The body of the letter should be approximately 180-200 words.

**Lancaster University**

**OET**  
OCCUPATIONAL ENGLISH TEST

**Lancaster University**

<https://www.occupationalenglishtest.org/preparation-portal/free-sample-tests/writing-sample-test-1/>

## Aspects of authenticity

TEST



'REAL WORLD'

1. Topics, texts and tasks reflect the TLU domain.
2. Tasks **prompt test-taker behaviour** that reflects behaviour in the TLU domain.
3. Criteria **capture features of test-taker performance** valued in the TLU domain.

## Weak and strong performance tests

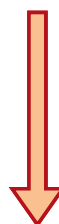
### Weak performance tests

*the focus is on language performance ...  
the purpose of the assessment is to elicit  
a language sample so that second language  
proficiency ... may be assessed*

### Strong performance tests

*performance will primarily be judged on  
real-world criteria, that is, the fulfilment of  
the task set*

(McNamara, 1996, pp. 43-44)



## Original speaking test criteria

Overall Communicative Effectiveness

Intelligibility

Fluency

Appropriateness of Language

Resources of Grammar and Expression



**OET**  
OCCUPATIONAL ENGLISH TEST



## Proposed speaking test criteria

### Language

### Clinical communication

#### Overall Communicative Effectiveness

#### Intelligibility

#### Fluency

#### Appropriateness of Language

#### Resources of Grammar and Expression



#### Clinician Engagement

- Professional manner
- Patient awareness

#### Management of Interaction

- Information-gathering
- Information-giving

(Elder et al., 2013; O'Hagan et al. 2014; Pill, 2016)

## Implemented speaking test criteria

### Linguistic criteria

OET SPEAKING Assessment Criteria and Level Descriptors (from September 2018) (public version)				
I. Linguistic Criteria				
Band	Intelligibility	Fluency	Appropriateness of Language	Resources of Grammar and Expression
6	• Pronunciation is easily understood and prosodic features (stress, intonation, rhythm) are used effectively • L1 accent has no effect on intelligibility	• Completely fluent speech at normal speed • Any hesitation is appropriate and not a sign of searching for words or structures	• Entirely appropriate register, tone and lexis for the context • No difficulty at all in explaining technical matters in lay terms • Confident use of idiomatic speech	• Rich and flexible • Wide range of grammar and vocabulary used accurately and fluently • Occasional errors in grammar or vocabulary are not intrusive
5	• Easily understood • Communication is not impeded by a few pronunciation or prosodic errors and/or noticeable L1 accent • Minimal strain for the listener	• Fluent speech at normal speed, with only occasional repetition or self-correction • Hesitation may occasionally indicate searching for words or structures, but is generally appropriate	• Mostly appropriate register, tone and lexis for the context • Occasional errors are not intrusive	• Wide range of grammar and vocabulary generally used accurately and fluently • Occasional errors in grammar or vocabulary are not intrusive
4	• Easily understood most of the time • Pronunciation or prosodic errors and/or L1 accent at times cause strain for the listener	• Spoken with some repetition, especially in longer utterances • Some evidence of searching for words, which does not cause serious strain • Delivery may be staccato or too fast/slow	• Generally appropriate register, tone and lexis for the context, but somewhat restricted and lacking in complexity • Errors are noticeable and at times limit limited resources of grammar and expression	• Sufficient resources to maintain the interaction • Inaccuracies in vocabulary and grammar, particularly in more complex sentences, are sometimes intrusive • Learning is generally clear
3	• Produce some acceptable features of spoken English • Difficult to understand because errors in pronunciation/prosodic features and/or L1 accent cause serious strain for the listener	• Very uneven • Frequent pauses and repetitions indicate searching for words or structures • Extended use of fillers and difficulty sustaining longer utterances cause serious strain for the listener	• Some evidence of appropriate register, tone and lexis, but errors are frequent and intrusive, reflecting inadequate resources of grammar and expression	• Limited vocabulary and control of grammatical structures, except very simple sentences • Persistent inaccuracies are intrusive
2	• Often unintelligible • Frequent errors in pronunciation/prosodic features and/or L1 accent cause serious strain for the listener	• Extremely uneven • Long pauses, numerous repetitions and self-corrections make speech difficult to follow	• Mostly inappropriate register, tone and lexis for the context	• Very limited resources of vocabulary and grammar, even in simple sentences • Numerous errors in word choice, structure
1	• Almost entirely unintelligible	• Impossible to follow, consisting of isolated words and phrases and self-corrections, separated by long pauses	• Entirely inappropriate register, tone and lexis for the context	• Limited in all respects
0	• Candidate does not provide any response			

### Clinical Communication Criteria

II. Clinical Communication Criteria	
In the roleplay, there is evidence of the test taker ...	
<b>A. Indicators of relationship building</b>	<b>A. Relationship building</b>
A1 initiating the interaction appropriately (greeting, introductions, nature of interview)	3 - Adopt use
A2 demonstrating an attentive and respectful attitude	2 - Competent use
A3 adopting a non-judgemental approach	1 - Partially effective use
A4 showing empathy for feelings/indications/emotional state	0 - Ineffective use
<b>B. Indicators of understanding &amp; incorporating the patient's perspective</b>	<b>B. Understanding &amp; incorporating the patient's perspective</b>
B1 eliciting and exploring the patient's ideas/concerns/expectations	3 - Adopt use
B2 picking up the patient's cues	2 - Competent use
B3 relating explanations to elicited ideas/concerns/expectations	1 - Partially effective use
<b>C. Indicators of providing structure</b>	<b>C. Providing structure</b>
C1 sequencing the interview purposefully and logically	3 - Adopt use
C2 signposting changes in topic	2 - Competent use
C3 using organising techniques in explanations	1 - Partially effective use
<b>D. Indicators for information gathering</b>	<b>D. Information gathering</b>
D1 facilitating the patient's narrative with active listening techniques, minimising interruption	3 - Adopt use
D2 using initially open questions, appropriately moving to closed questions	2 - Competent use
D3 NOT using compound questions/leading questions	1 - Partially effective use
D4 clarifying statements which are vague or need amplification	0 - Ineffective use
D5 summarising information to encourage/encourage further information	
<b>E. Indicators for information giving</b>	<b>E. Information giving</b>
E1 establishing initially what the patient already knows	3 - Adopt use
E2 pausing periodically when giving information, using the response to guide next steps	2 - Competent use
E3 encouraging the patient to contribute reactions/feelings	1 - Partially effective use
E4 checking whether the patient has understood information	0 - Ineffective use
E5 discovering what further information the patient needs	

## Authenticity of criteria

### Examples

- Cambridge First speaking test – **interactive communication** criterion
- Research into criteria for a test of English as a Lingua Franca

Competence area	Description
Accommodation	Making self intelligible to interlocutor
Negotiation	Adjusting to interlocutor's speech style/accent Clarification Self-repair Repetition Paraphrasing
Maintaining smooth interaction	Turn-taking Using appropriate politeness strategies

(Harding & McNamara, 2017)

## Some final considerations

- Inauthenticity of separating tests into four skills – RLWS
- Can tests ever be completely authentic to real life?  
We are socialised to know what to expect of a test
- In any case, aren't tests themselves part of the real world?  
Tests have their own authenticity

***Authentic tests of language* vs *Tests of authentic language***

(Lewkowicz, 2000)



## References

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Elder, C., McNamara, T., Woodward-Kron, R., Manias, E., McColl, G., Webb, G., & Pill, J. (2013, August). *Towards improved healthcare communication: Development and validation of language proficiency standards for non-native English speaking health professionals. Final report for the OET Centre*. (40 pp.) Melbourne, Australia: University of Melbourne.
- Harding, L., & McNamara, T. (2017). Language testing: The challenge of ELF. In J. Jenkins, W. Baker & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 570-582). London, UK: Routledge.
- Knoch, U., & Macqueen, S. (2016). Language assessment for the workplace. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 291-307). Berlin, Germany: De Gruyter.
- Lewkowicz, J. (2000). Authenticity in language testing: Some outstanding questions. *Language Testing*, 17(1), 43-64.
- McNamara, T. F. (1996). *Measuring second language performance*. London, UK: Longman.
- O'Hagan, S., Manias, E., Elder, C., Pill, J., Woodward-Kron, R., McNamara, T., Webb, G., & McColl, G. (2014). What counts as effective communication in nursing? Evidence from nurse educators' and clinicians' feedback on nurse interactions with simulated patients. *Journal of Advanced Nursing*, 70(6), 1344-1354.
- Pill, J. (2016). Drawing on indigenous criteria for more authentic assessment in a specific-purpose language test: Health professionals interacting with patients. *Language Testing*, 33(2), 175-193.