Are current academic reading tests fit for purpose?

Cyril Weir and Sathena Chan

Construct construct construct

Alan Davies wrote in the first issue of the journal *Language Testing*:

"...in the end no empirical study can improve a test's validity... What is most important is the preliminary thinking and the preliminary analysis as to the nature of the language learning we aim to capture."

Davies, A. (1984). Validating three tests of English language proficiency. *Language Testing*, *1* (1), 50-69.

The EAP reading construct

Some examples of the empirical research data we drew on to establish the parameters that might be included in a **model of reading for academic purposes**:

- Weir (1983) who investigated the language activities and associated problems of students studying at tertiary level throughout the UK in the early 1980's; Responses were received from 940 overseas students, 530 British students and 560 academic staff, in respect of 43 post graduate courses, 61 undergraduate courses and 39 'A' level courses

- Weir, Yang and Jin (2000) who carried out a similar analysis of academic English needs of first year undergraduates in China;

- Weir, Hawkey, Green, Unaldi and Devi (2007/12) who carried out a needs analysis of undergraduates' and postgraduates' language needs and problems at the University of Bedfordshire in the 21st century. 766 students responded.

Research Study 1
Weir (1983) on academic reading

In summary, Weir's (1983) study:

- *reveals the importance of assessing both careful and expeditious global reading in a test of academic English;*

- *makes clear the need for students to process information at both the whole text and intertextual levels, rather than just comprehending at the paragraph or sentence levels*

Research Study 2

Weir, Huang and Yan's (2000) study recommends:

- *Texts used in academic English tests should mirror the contextual variables found in real life academic texts*

- *Expeditious reading must be tested as well as careful reading in academic English tests*

- *Separability of reading skills should be mirrored in the way results are reported*

- *If a choice is to be made in what is tested, global items i.e. comprehension beyond the sentence must always take precedence over local items.*

Research Study 3

For those intending to develop EAP reading tests, Weir et al's (2012a) study provides clear evidence for the necessity of:

- *a dedicated expeditious global reading task to assess search reading of a longer text under time constraints*

- *items testing the higher-level skills of understanding main ideas,*

- *items testing how a text is structured*

- *items testing the ability to integrate information across different texts*

6

"ACADEMIC READING" TESTS

i) the contextual parameters in academic reading tests

ii) the types of reading required for measuring the construct of academic reading

iii) the levels of processing in reading which should be prioritised in tests of academic reading,

iv) formats and items in tests of academic reading

v) assessing higher-level academic reading through an integrated task type,

i) **Contextual parameters**

There are a number of quantifiable differences between the texts that students are exposed to in their first year of study and those used in IELTS reading passages.

IELTS texts do appear in some respects to be "easier" than the UG texts students are faced with. The main areas of concern in our data are that the IELTS texts:

1. **generally include a lower proportion of sub-technical academic vocabulary than the undergraduate texts**
2. **IELTS texts may not fully reflect the level of readability found among the more challenging academic texts that first year undergraduates might expect to encounter.**

Lexical/syntactic complexity

Length C1000 words IELTS average 854: min 589- max1063, CAE max 1100, GEPT ave 700, TOEFL ave 700 (**Density and complexity of idea units** important here)

Type Token Ratio (> 55%) CAE 59. 95 TOEFL 55 5 GEPT 47.74 IELTS 53.68 UG 51.7

Lexical density (> 55%) CAE O.52 TOEFL 0.59 GEPT 0.59 IELTS ,O.57 UG 0.56

Mean of modifiers per noun phrase (0.95-1.0) CAE 0.92 TOEFL 1. GEPT 0.98 IELTS 0.98 UG 0.95

Mean number of words before main verb in main clause (5.5) CAE 4.4 TOEFL 5.7 GEP T 5.52, IELTS 5.48 UG 4.59

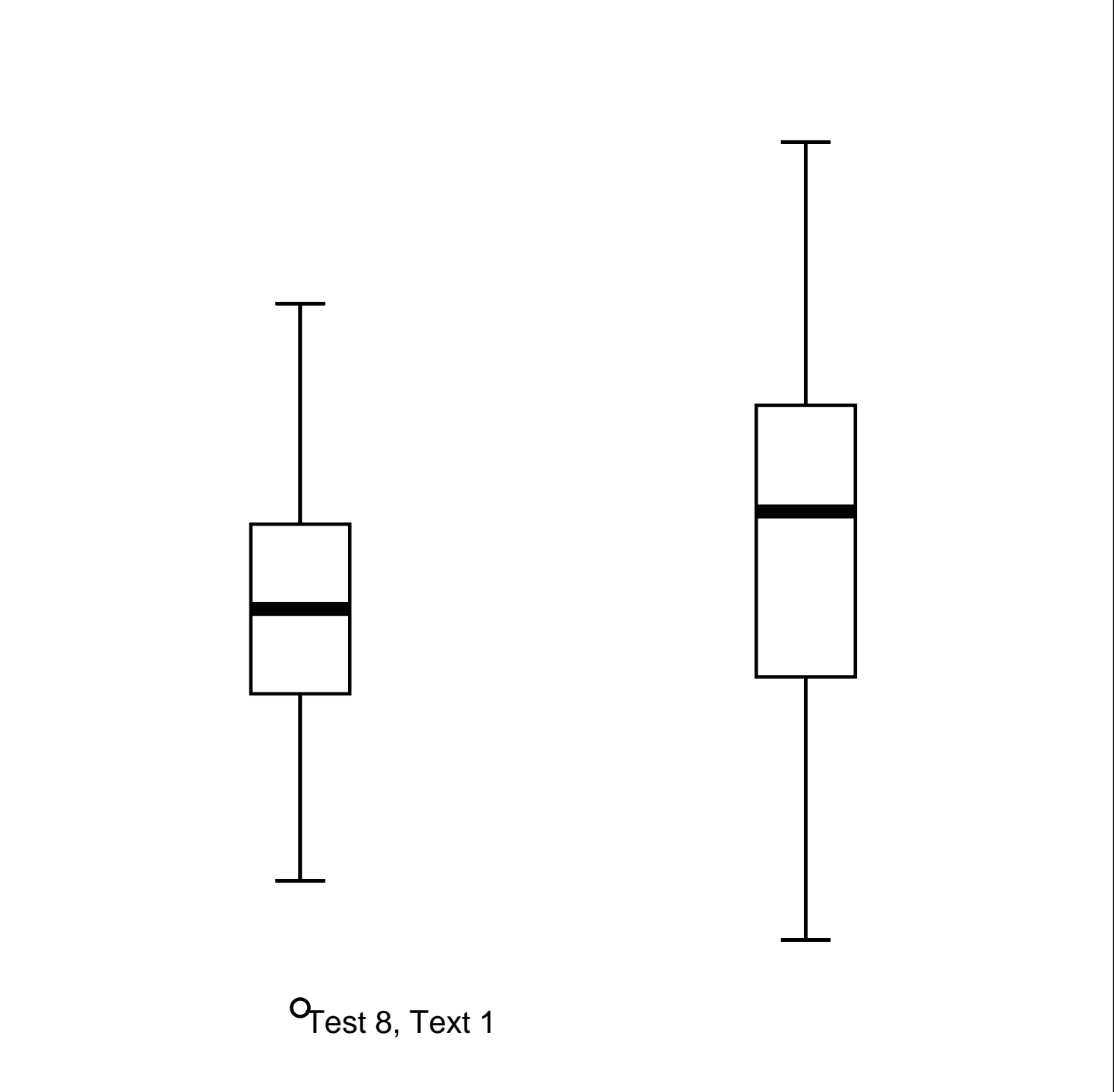***Academic Word List >10%** IELTS.7.9, iBT TOEFL 7.84, CAE 4.58, GEPT 7.0 , UG 10.5 %

Grammatical level a range of sentence patterns, frequent compound sentences and **many complex sentences**

Average sentence length 20-22 IELTS average 21.89, CAE 20.42, TOEFL 19.26, GEPT Advanced 21, UG 21.47

*Readability Flesch-Kincaid Grade Level of **=/>13.5** IELTS average FK 12.64 TOEFL 12.07 UG 13.5, GEPT 12 CAE 9.6

Test 8, Text 1

The IELTS compromise: One size fits all?

An issue to resolve is whether current or any revised academic reading tests can be used appropriately in a variety of domains or whether in the future, variants may need to be offered for example in professional as against academic situations

Appropriate for professions?

Read and Wette (2009) investigated the attitudes of a group of overseas-trained health professionals towards the IELTS test in their experiences of meeting the English language requirements for professional registration. Their study found that participants were generally favourably disposed towards the test being used for this purpose, while at the same time recognising that IELTS "... had limited capacity to assess their ability to communicate effectively in clinical settings" (3) Read and Wette note in relation to this response that IELTS is still designed primarily as a test for those entering educational and training contexts, and is not specifically intended "to assess the communication skills required in particular professions" (p 4).

Appropriate for low level learners?

- As well as concerns for its suitability for assessing medical staff and the professions, questions have also been raised, for example, about its suitability for assessing lower level EAP students in the Middle East. Morrow (2017:1) asks:

- "While the IELTS exam has earned a reputation for its usefulness in many international educational contexts, it does not appear to fully address the specific needs and issues of students and universities in the United Arab Emirates… its influence on instructional and assessment practices has grown … in ways that do not seem to be completely congruent with the needs of low-level English learners who are about to begin studies in English-medium universities."

Readiness to enter: time for a change?

The concept of 'readiness to enter' has been used as an argument for current IELTS to be like it is i.e. general and not based squarely on the academic domain or on professional practice, a test for all seasons as it were rather than fit for one particular purpose.

Do we need a general test of a candidate's baseline competence, at say B2/C1 levels in the four skills which predicts <u>readiness</u> to enter professional or academic situations (rather like Aptis in fact) or do we require a test of whether the candidate can actually cope with the demands initially faced in a profession or in an academic context?

If it is the former, little change is needed. If it is the latter, then clearly more than one 'authentic' needs driven test will be required and the logistical challenges to be faced that much greater.

TOEFL: an academic English test

The TOELF iBT, which is radically different from earlier versions of this test, clearly definines itself as a measure of academic language ability and aligns itself more closely to "the variety of language use tasks that examinees are expected to encounter in everyday academic life' (Sawaki et al., 2009, p. 5).

Sawaki, Y.; Stricker, L. J.; Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. Language Testing, 26 (1), 5–30.

What is an academic reading test?

He who loves practice without theory is like the sailor who boards ship without a rudder and compass and never knows where he may cast.

Leonardo da Vinci *The Notebooks of Leonardo Da Vinci*, translated by Jean Paul Richter (1888), I - Prolegomena

ii) **Types of reading required**

Reading Type	Global level	Local level
Careful Reading	- Establishing accurate comprehension of explicitly stated main ideas and supporting details across sentences - Making propositional and/or pragmatic inferences - Establishing how ideas and details relate to each other in a whole text - Establishing how ideas and details relate to each other across texts	- Establishing accurate comprehension of explicitly stated main idea or supporting details within a sentence - Identifying lexis - Understanding syntax
Expeditious Reading	- Skimming quickly to establish: discourse topic and main ideas, or structure of text, or relevance to needs - Search reading to locate quickly and understand information relevant to predetermined needs	Scanning to locate specific points of information

Expeditious reading

'He has only half learnt the art of reading who has not added to it the even more refined accomplishments of skipping and of skim-ming.'

Arthur Balfour "*The pleasures of reading*" (**1905**: 32)

Careful v Expeditious Reading

Weir et al (2013) point out that though a great deal of attention has been paid to the assessment of slow, careful, incremental reading, expeditious reading (fast, efficient, selective reading) has not been explicitly tested in very many high stakes examinations to this day, despite its inclusion in CEFR reading descriptors for academic purposes at the C1 levels (Council of Europe, 2001).

The 1991 - Cambridge CAE reading paper was a notable exception to this, as were the later multiple matching tasks in the PET and FCE.

Separate expeditious paper

The Higher Intermediate General English Proficiency Test (GEPT) in Taiwan has had a separate expeditious reading paper since its inception in 2000 (Weir 2005a, Wu & Lin 2008, Wu 2013) but apart from GEPT and the Test for English Majors (TEM) in China (See Zou, Green & Weir, 1997) these appear to be the only high stakes English language examinations which have papers dedicated to expeditious reading.

Need an expeditious reading task

- **Weir et al (2012b: 179) conclude that there should be a dedicated section in any academic reading test to assess expeditious global reading. Such a test would impose explicit time constraints that encourage the use of the expeditious global reading strategies essential for university study.**

- **If the test is computerised then such an expeditious task is clearly both possible and desirable.**

- **Such dedicated tasks are likely to have beneficial washback if they equip students with the strategies to read texts selectively, quickly and efficiently to establish global meaning, as they have to do in real life reading activities in an academic context.**
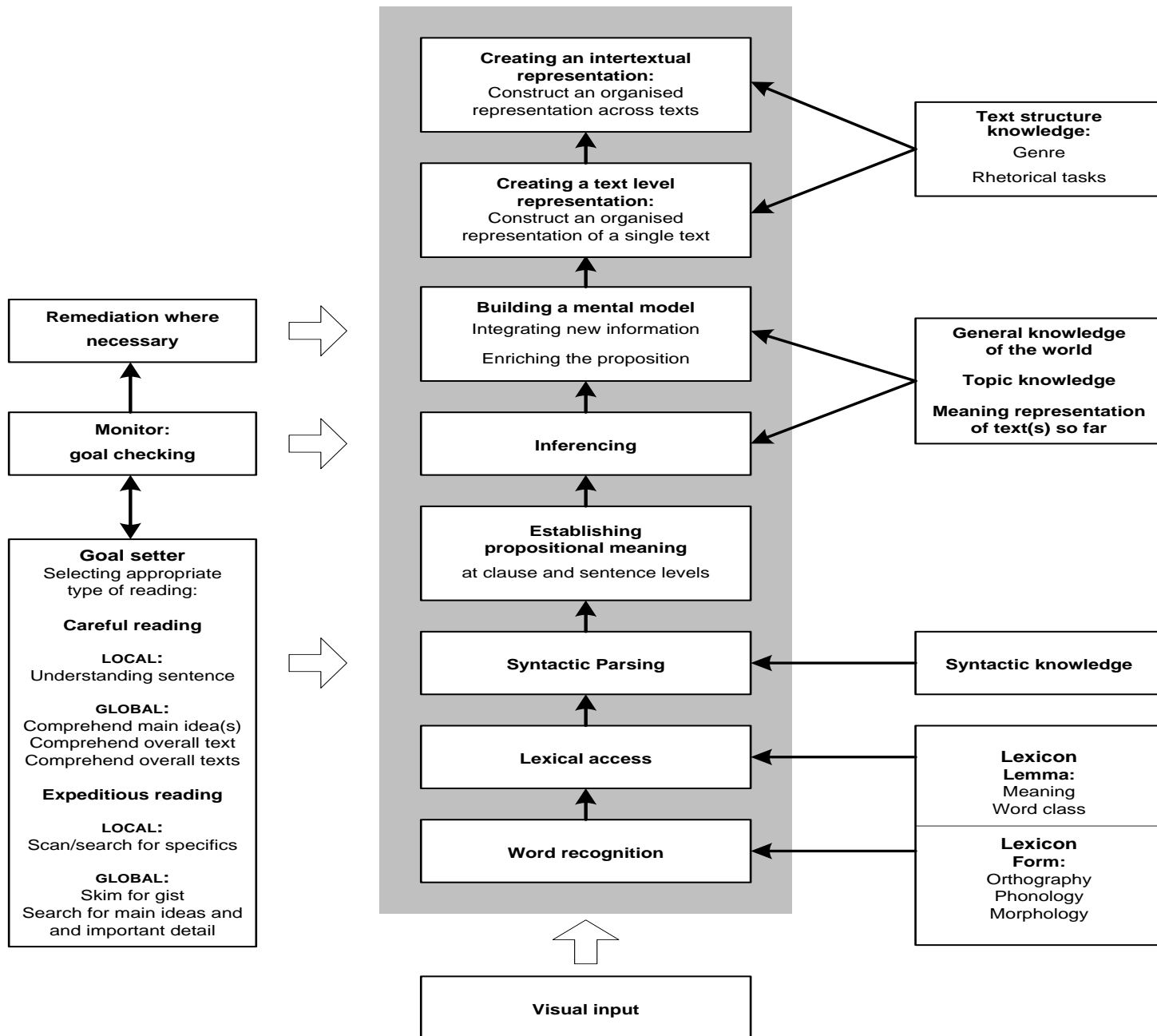
iii) The levels of processing in reading which should be prioritised in tests of academic reading

The research literature on reading in an academic context, and on IELTS reading in particular (see for example Moore, Morton and Price, (2007), Green and Hawkey, (2012), Weir et al, (2012a and 2012b), Chalmers and Walkinshaw (2014, p.24), Bax (2015), and Owen (2016), are unanimous that academic reading tests should involve higher order reading processes rather than lower order processes.

Owen 2016 Doctoral Thesis

Owen concludes in his findings on IELTS and TOEFL (2016: 366-367):

The majority of identified cognitive processes were lower level processes for both tests. ... high level processing in both tests lies primarily in 'forming a mental model' (P6) rather than inferencing or creating a text-level representation... Some of the higher-level processes in Khalifa and Weir's model are under-represented in both IELTS and TOEFL, including inferential reasoning and forming a text-level representation

Global versus local reading

- By the twenty first century the IELTS reading test appears to be for the most part a test of general reading ability with items focusing mainly on the <u>local</u> level.

- If we examine the cognitive focuses of the current IELTS academic reading we can see how far the shift from global to local reading items has gone.

- Given the clear need for first year university students to understand the main ideas in a text and how these relate to each other within and across texts in writing their assignments, this preponderance of items focusing on lower level reading skills must be a cause for some concern.

- There is a clear need to ensure that future IELTS tests at the global level include items that require creation of (i) a text level representation and (ii) an intertextual representation.

iv) formats and items in tests of academic reading

- The prevalence of items testing at the lower levels of reading in current IELTS is indicative that a historical shift in the focuses of the test has occurred, a direction compounded by the item formats currently used.

- Serious thought should be given to reducing the number of items/tasks that appear to initiate only lower order processing to ensure IELTS is only testing global meaning beyond the sentence; a sine qua non of academic reading.

- The current formats testing lower order reading processes need to be replaced by item types that are likely to require higher order reading processes particularly item types that lead to a textual or inter-textual representation.

Alderson on the fallacy underlying current reading tests

Having described the most common formats used for testing reading comprehension in great detail over c.50 pages (Alderson, 2000, p. 202-248), Alderson then raises the uncomfortable question of whether any of the formats he has suggested could be said to represent what we actually do in reading in real life i.e.:

> "correspond to the way they normally read such texts…the test may not reflect how students would understand the text in the real world…the only purpose we typically give students for their reading is to answer our questions, to demonstrate their understanding or lack of it… readers do not usually answer somebody else's questions: they generate and answer their own" (Alderson 2000: 249)

v Assessing higher-level academic reading through an integrated task type

The thinness of the soil is often displayed by the English Essay and frequently its prognostication proves correct... (276)

I would allow candidates while writing their essay in the examination room to have access to some standard authorities on their subject... there is now no object in forcing men to carry a number of details in their heads...the range of subjects which can be given for essays is very much extended

Henry Latham *On the Action of Examinations Considered as a Means of Selection* 1877

Beyond reading for comprehension

- Critical thinking
  - synthesize multiple ideas into a theory
- Reading
  - determine major and subordinate ideas in passages
  - summarize information
  - synthesize information from reading and incorporate it into a writing assignment
  - read texts of complexity without instruction and guidance
  - retain information while searching for answers to self-generated questions

Intersegmental Committee of the Academic Senates of the California Community Colleges and the University of California (2002) *Academic Literacy: A Statement of Competencies Expected of Students Entering California's Public Colleges and Universities.* Available at www. http://icas-ca.org/acadlit.

Which academic reading skills can be tested better by the R/W format?

In our view, three of the 14 academic reading skills identified in Weir (1983) could be tested better by the integrated format.

- Separating the essential from the non-essential in a text
- Note-making:

    (a) Extracting salient points for summary

    (b) Selective extraction of relevant points from a text

    (c) Reducing a text through rejection of redundant information

- Critical evaluation: assessing the worth of a text and the way information in it has been organised and expressed

Evidence from the literature

- When students are required to read for writing, they engage in the more complex, higher order processes of constructing textual and intertextual representations when compared to responding simply to conventional reading comprehension items (Weigle, Yang, and Montee, 2013)

- When reading to integrate materials, students construct elaborate models of the text structure and situation, enabling them to select information from the source text, evaluate it, and use it for writing purposes (Kintsch, 1998).

- The various goals of reading-into-writing elicit different higher-level reading strategies, e.g. more integrative and analytic tasks leading to richer representation of the texts (Langer, 1980, 1986; Newell, 1984)..

- Reading ability was found to be linked to characteristics of students' syntheses of sources, especially organisation (Spivey, 1988).

The advantages of using the integrated format to test academic reading skills

A reading-into-writing task

- gives students a reader's perspective (a goal for comprehending with a communicative intent of its own) – which in turn affects the attention allocated to different parts of the texts (Goetz, Schallert, Reynolds, & Radin, 1983; Just & Carpenter, 1980)

- allows/encourages students' autonomy to handle source materials rather than responding to pre-constructed representations

- elicits deeper and more academically authentic reading processes as students interpret the full significance of the source material (Weigle, Yand and Montee, 2013)

Reading into writing

- Reading into writing tasks represent the most situationally and interactionally authentic test task for testing academic reading ability (and writing) i.e. they are the most construct valid tasks for this purpose.

- It may well be time to revisit the use of integrated reading into writing tasks to better represent the higher order processing in reading that is the hallmark of academic reading (and higher order processing in writing).

- It is perfectly plausible to score the reading element of such tasks according to the **relevance** and adequacy of the content produced, the way **connections** have been made between ideas and **organised**, though the availability of a construct valid piece of writing would be an added bonus and bring other criteria into play if it was to be included as part of the writing assessment

Are current academic reading tests fit for purpose?

Can we develop fit for purpose academic reading tests?

YES, WE CAN.

THE END