7/12/2013

University of Bedfordshire

Business School

Validating two types of EAP reading-into-writing test tasks

Sathena H C Chan

Validating two types of EAP reading-into-writing test tasks

▸ Weir's (2005) socio-cognitive validation framework has been widely used in test validation research, but its current application is limited to independent language tests.

▸ The study investigated the construct validity of two operationalised EAP reading-into-writing test tasks in terms of:
   a) Context validity – the task features
   b) Cognitive validity – the cognitive processes elicited by the tasks
   c) Predictive validity – relationship between test scores and real-life performance

▸ 2

The two real-life tasks and the two reading-into-writing test tasks

▸ Two real-life academic writing tasks were selected from 8 modules

	Real-life Task A	Real-life Task B
Genre	Essay	Report
Input format	Multiple verbal and non verbal	Multiple verbal and non verbal

▸ Two operationalised reading-into-writing test tasks were selected

	Test Task A	Test Task B
Function	Criterion-referenced level specific test	University's diagnostic test
Level	C1	B2
Input format	2 articles without non-verbal input	2 articles with a non-verbal input each

▸ 3

Investigating the cognitive validity of EAP reading-into-writing test tasks

Research methods (built on Green et al, 2010, 2012; Weir, 2102; Wu, 2012)

▸ Expert judgement + automated textual analysis

Participants

▸ An expert panel of 10 judges

Instruments

▸ Contextual parameter proforma (built on Shaw and Weir, 2007; Wu, 2012)

  ▸ Overall task setting
  ▸ Features of the input texts

▸ Automated textual analysis tools

  ▸ Coh-Metrix version 2.1 (Graesser, McNamara, Louwerse and Cai, 2004)
  ▸ VocabProfile version 3 (Cobb, 2003)

▸ 4

Contextual Parameter Proforma

Part 1 - Overall task setting					
1. Purpose	1 Unclear	2	3	4	5 Clear
2. Topic Domain (Please circle a rating for each domain)	Personal Not at all ... Definitely 1 2 3 4 5	Social Not at all ... Definitely 1 2 3 4 5	Academic Not at all ... Definitely 1 2 3 4 5	Professional Not at all ... Definitely 1 2 3 4 5	
3. Genre	Essay	Report	Case Study	Summary	Others (Please specify):
4. Cognitive demands	1. Telling personal experience / viewpoints		2. Summarising / organising given ideas		3. Transforming given ideas into new representations
5. Language functions to perform (you may choose more than 1)	Classify Persuade Synthesise (to combine different (parts of) texts to form a new text with own interpretations)	Cite sources Predict Express personal views	Describe Recommend Illustrate visuals	Define Reason Others (Please specify):	Evaluate Summarise
6. Intended reader	1 Unclear	2	3	4	5 Clear
7. Knowledge of criteria	1 Unclear	2	3	4	5 Clear

5

Contextual Parameter Proforma

Part 2 - Input text features					
8. Input format	Single verbal	Single non-verbal	Multiple verbal	Multiple non-verbal	Multiple verbal and multiple non-verbal
	Others (Please specify):				
9. Verbal input genre	Book chapter	Journal article	News / Magazine article	Proposal	Report	Review
	Others (Please specify):				
10. Non-verbal input	Table	Graph	Diagram	Picture	
11. Discourse mode (Consider the primary purpose of the text)	Narrative	Descriptive	Expeditionary	Argumentative	
12. Concreteness of ideas	1 Abstract	2	3	4	5 Concrete
13. Textual organisation	1 Inexplicit	2	3	4	5 Explicit
14. Cultural specificity	1 Neutral	2	3	4	5 Specific

6

Automated textual analysis tools

▶ Automated textual analysis tools

  ▶ The usefulness of the all Coh-Metrix and VocabProfile indices (built on Green et al, 2010, 2012; Weir, 2102; Wu, 2012) was evaluated. 30% of the real-life input texts were analysed in the pilot

  ▶ **13** Coh-Metrix and **4** VocabProfile indices were selected to compare the difficulty level between real-life input texts and reading-into-writing test task input texts in terms of: (1) lexical complexity, (2) syntactic complexity and (3) degree of cohesion.

Real-life input texts	Reading-into-writing input texts	Analysis
60 extracts from 20 texts	Test Task A: 20 passages	Inferential statistics
60 extracts from 20 texts	Test Task B: 2 passages	Descriptive comparison

▶ 7

Selected automated textual indices

Lexical complexity	Syntactic complexity	Degree of cohesion
• High frequency words (K1) • High frequency words (K2) • Academic words • Low frequency words (Offlist) • Log frequent content words • Average syllables per word • Type-token ratio (content words)	• Average words per sentence • Sentence syntax similarity • Mean number of modifiers per noun-phrase • Mean number of words before the main verb • Logical operator incidence	• Adjacent overlap argument • Adjacent overlap stem • Adjacent overlap content word • Proportion of adjacent anaphor references • Adjacent semantic similarity (LSA)

U1

▶ 8

U1   Lexical = decoding
Syntactic = higher-level processing
the more complex sentence structures a text contains, the more difficult it is for readers to process the text

degree of coherence = main theme

The more cohesive a text is, the easier it would be for the reader to build the textual representation
USER, 10/07/2013

Investigating the cognitive validity of EAP reading-into-writing test tasks

▸ Participants: 219 undergraduates

▸ Research instruments:

Writing phases (Field, 2004; 2013; Shaw and Weir, 2007)	No. of items
Conceptualistion	8
Meaning and discourse construction	11
Organisation	9
Low-level monitoring and revising	8
High-level monitoring and revising	12

- A writing process questionnaire was developed based upon models of reading (Khalifa and Weir, 2009), writing (e.g. Hayes and Flowers, 1983; Kellogg, 1994, 1996) and discourse synthesis (Spivey, 1990, 1997).

▸ 9

Investigating the cognitive validity of EAP reading-into-writing test tasks

▸ Data collection

Conditions	Tasks	N	Total
Real-life	A (Essay)	70	
	B (Report)	73	143
Test	A (multiple verbal input)	160 (81 did both + 79 did only A)	
	B (multiple verbal and non-verbal input)	140 (81 did both + 59 did only B)	300
		Total	443

▸ 10

Investigating the cognitive validity of EAP reading-into-writing test tasks

▸ Data analysis

Investigating the real-life cognitive constructs

- Descriptive statistics of individual questionnaire items from each of the real-life tasks
- Comparison of the processes elicited by the two real-life tasks
- Comparison of the processes employed by high-achieving and low-achieving participants
- Exploratory factor analysis of the underlying structure of the cognitive processes involved in the five academic writing phases

Investigating the cognitive processes elicited by the test tasks

- Descriptive statistics of individual questionnaire items from each of the reading-into-writing test tasks
- Comparison of the processes elicited by a) the two real-life tasks and Test Task A, and b) the two real-life tasks and Test Task B
- Exploratory factor analysis of the underlying structure of the cognitive processes involved in the five academic writing phases elicited by a) Test Task A and b) Test Task B

▸ 11

Investigating the predictive power of the two reading-into-writing test tasks

▸ Four points of reference were selected

- Essay
- Report
- In-class question-and answer test
- End-of-term case study exam

Condition	Tasks	No of scores collected
Real-life academic context	Essay	161
	Report	136
	In-class question and answer test	145
	End-of-term case study examination	143
Reading-into-writing language tests	Test Task A	160
	Test Task B	140

▸ 12

Findings of the contextual features of real-life academic writing tasks and the context validity of the reading-into-writing test tasks

▶ 13

Findings of context validity – Overall task setting

1. Genre
   - Real-life Task A **(Essay)**; Real-life Task B **(Report)**
   - Test Task A **(Essay)**; Test Task B **(Essay)**
   - 'Although the test tasks both require the test takers to write "an essay", both tasks require the test takers to write a summary in a more specific term. Essay can be anything' (Judges Pair 1).

2. Clarity of purpose (1=unclear;5=clear)
   - It is interesting that Test Task A and Test Task B seemed to present a clearer purpose than the real-life tasks did.
   - Real-life Essay **(3.6)**; Real-life Report **(4.4)**
   - Test Task A **(4.6)**; Test Task B **(4.8)**
   - 'There was hardly a real communicative purpose to achieve on this task [real-life essay], apart from following the instructions' (Judges Pair 2)

▶ 14

Findings of context validity –
Overall task setting

### 3. Topic domain (Clapham, 1996; Douglas, 2000; Read, 1990)

- Real-life Essay **(professional** and academic**)**
- Real-life Report **(academic** and professional**)**
- Test Task A **(academic** and social**)**
- Test Task B **(professional** and social**)**
- The judges felt that both test tasks' input texts contained rather general content, which was usually connected to the social domain.
- Issues to consider: a continuum of content specificity

▶ 15

Findings of context validity –
Overall task setting

4. Cognitive demands (Galbraith and Torrance, 1999; Purves et al, 1984; Scardamalia and Bereiter, 1987)

1) **Telling / retelling content** - a linear 'think-write' or what next?' process
2) **Organising / reorganising content** – the writer needs develop an explicit representation of the rhetorical problem of the writing task and purposefully organise the content they retrieved from long-term memory and/or selected from the input texts in order to solve the rhetorical problem of the writing task.
3) **Transforming content** – a process requires the writers a contribution of transformed or new knowledge through the activation of high-level processes, such as *integration, interpretation, elaboration, evaluation* and *modification*.
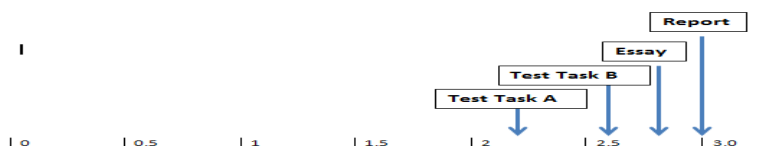
▶ 16

Findings of context validity –
## Overall task setting

5. Language functions to perform

- Core language functions, those were judged by 2 or more pairs of the judges, required by the two real-life tasks included *describing, defining, reasoning, citing sources, evaluating, synthesising* and *expressing personal views.*

Real-life essay	Real-life report	Test Task A	Test Task B
• Reasoning • Express personal views • Cite sources • Evaluate • Persuade • Synthesise • Describe • Summarise • Define	• Describe • Define • Reasoning • Illustrate visuals • Cite sources • Evaluate • Predict • Recommend • Synthesise • Express personal views	• Summarise • Express personal views • Cite sources • Evaluate • Recommend • Reasoning • Synthesise • Describe	• Reasoning • Summarise • Express personal views • Evaluate • Recommend • Synthesise • Illustrate visuals

17

Findings of context validity –
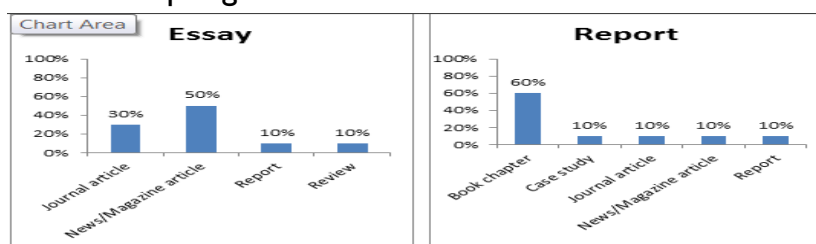## input text features
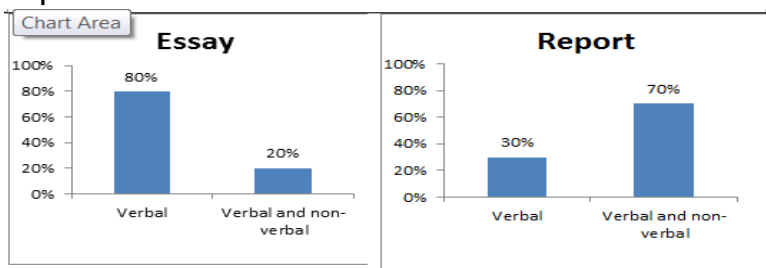
8. Input format

- The input format of the two reading-into-writing test tasks is standardised. Test Task A contains **two reading passages** while Test Task B contains **two passages with a non-verbal input** in each.

18

Findings of context validity –
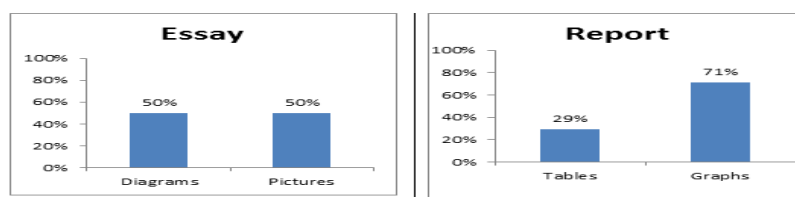## Input text features

9. Verbal input genre

- Test Task A: a simplified version of the **essay** genre
- Test Task B: **news/magazine article** and **report**.

19

Findings of context validity –
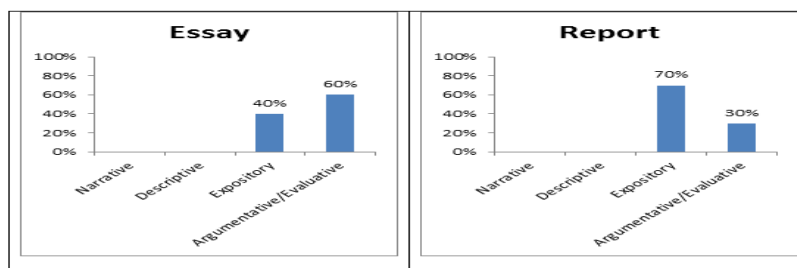## Input text features

10. Non-verbal input type

- Test Task A contained no non-verbal input.
- Test Task B contained two verbal inputs - both were diagrams.

20

Findings of context validity –
Input text features

11. Discourse mode

- The input texts on the test tasks were dominated by single discourse mode.
- All texts in Test Task A were identified as **argumentative** texts while all texts in Test Task B contained only **expository** texts.

21

Findings of context validity –
Input text features

12. Concreteness of the ideas (1=abstract; 5=concrete)
  - The ideas in the test task input texts were considered more concrete than those in the real-life input texts.
  - Real-life Essay **(3.25)**; Real-life Report **(2.9)**
  - Test Task A **(4.1)**; Test Task B **(4)**

13. Explicitness of the textual organisation (1=Inexplicit; 5=Explicit)
  - The judges felt that some of the test task texts were too explicitly organised by using rather formulaic markers such as 'firstly', 'in addition', 'lastly', etc.
  - Real-life Essay **(3.5)**; Real-life Report **(3.15)**
  - Test Task A **(3.9)**; Test Task B **(4)**

22

Findings of the cognitive processes elicited by real-life academic writing tasks and the cognitive validity of the reading-into-writing test tasks

23

Findings of cognitive validity –
Defining the **real-life** academic writing processes

▸ Based on the results of explanatory factor analysis, the hypothesised academic writing phases arising from the literature review were largely supported by the statistical analysis of the questionnaire data collected in this study.

Academic writing phases	Cognitive processes
Conceptualisation	F1: Task representation and macro-planning F2: Revising macro plan
Meaning and discourse construction	F1: Connecting and generating F2: Selecting relevant ideas F3: Careful global reading
Organisation	F1: Organising ideas in relation to input texts F2: Organising ideas in relation to own text
Low-level monitoring and revising	F1: Low-level editing after writing F2: Low-level editing during writing
High-level monitoring and revising	F1: High-level editing after writing F2: High-level editing during writing

24

Findings of cognitive validity –
Defining the **real-life** academic writing processes

- The processes employed by high-achieving and low-achieving participants on real-life tasks
  - Based on the results of Mann-Whitney U tests, the high achieving participants reported employing eight of the eleven cognitive processes (i.e. *task representation and macro-planning, careful global reading, selecting relevant ideas, connect and generate, organising ideas in relation to source texts, organising ideas in relation to new text, low-level editing while writing* and *high-level editing while writing*) more than the low achieving groups**.**

25

Findings of cognitive validity –
The cognitive processes elicited by the test tasks

- **Comparison between the cognitive processes elicited under test conditions and the real-life conditions (in groups of high-, mid- and low-achievement)**
  - Based on the results of the Wilcoxon signed ranks tests, both reading-into-writing test tasks were able to elicit from high-achieving and low-achieving participants most of the cognitive processes to a similar extent as participants employed the processes on the real-life tasks.
  - The middle group showed greater discrepancy in how they employed the processes under the test and real-life conditions. Generally speaking, they tended to employ some processes more in the real-life conditions than the test conditions. They employed the processes of the *task representation and macro-planning, revising macro-plan, low-level monitoring and revising,* and *high-level monitoring and revising* phases significantly less on the test tasks than on real-life tasks. In addition, there seemed to be an over-eliciting of *careful global reading* more on Test Task B.

26

Findings of cognitive validity –
The cognitive processes elicited by the test tasks

▸ Explanatory factor analysis to examine the underlying structure of the processes within each writing phase elicited by Test Task A and Test Task B

- The underlying structures of the four out of five phases of academic writing, which include **conceptualisation**, **organising**, **low-level organising and revising**, and **high-level monitoring and revising**, elicited on Test Task A and the real-life tasks were identical.

- The underlying structures of the cognitive processes of four phases of academic writing, which include **discourse and meaning construction**, **organising**, **low-level organising and revising**, and **high-level monitoring and revising**, elicited on Test Task B and the real-life tasks were identical.

▸ 27

The underlying structure of the processes involved in each writing phase

	Real-life tasks	Test Task A	Test Task B
	Conceptualisation phase		
F1:	Task representation and macro-planning (34%)	Task representation and macro-planning (33.98%)	Macro-planning (31.98%)
F2:	Revising macro plan (19.9%)	Revising macro plan (19.04%)	Revising macro plan (16.16%)
F3:			Rereading task prompt (16.16%)
	Meaning and discourse construction phase		
F1:	Connect and generate (34.54%)	Selecting relevant ideas (33.38%)	Selecting relevant ideas (28.20%)
F2:	Selecting relevant ideas (13.88%)	Connect and generate with careful global reading (14.09%)	Connect and generate (18.45%)
F3:	Careful global reading (10.16%)	-	Careful global reading (15.07%)

▸ 28

The underlying structure of the processes involved in each writing phase

Organising phase		
F1: Organising ideas in relation to input texts (34.73%)	Organising ideas in relation to input texts (41.70%)	Organising ideas in relation to input texts (40%)
F2: Organising ideas in relation to own text (16.60%)	Organising ideas in relation to own text (17.65%)	Organising ideas in relation to own text (16.41%)
Low-level monitoring and revising phase		
F1: Low-level editing after writing (47.70%)	Low-level editing after writing (48.18%)	Low-level editing after writing (52.31%)
F2: Low-level editing while writing (23.9%)	Low-level editing while writing (19.31%)	Low-level editing while writing (24.85%)
High-level monitoring and revising phase		
F1: High-level editing after writing (42.92%)	High-level editing after writing (47.42%)	High-level editing after writing (47.14%)
F2: High-level editing while writing (24.35%)	High-level editing while writing (19.46%)	High-level editing while writing (23.51%)

29

Findings of cognitive validity –
The cognitive processes elicited by the test tasks

▸ **Seven factors** within these phases elicited by **Test Task A,** and **eight factors** by **Test Task B** contained the same individual questionnaire items as the corresponding factors identified by the real-life tasks.

▸ Processes that might require further attention
- The processes of careful global reading did not load as an independent factor on Test Task A's data (essay with verbal and non-verbal inputs).
- There is a seemingly over-eliciting of careful global reading on Test Task B (essay with verbal and non-verbal inputs).
- The factor of *organising ideas in relation to own text* elicited on both test tasks involved less items than the corresponding factor identified by real-life data.

30

Findings of the predictive power of the reading-into-writing test scores

▶ 31

Findings of criterion-related validity –
Predictive power of the two reading-into-writing test tasks

▶ Of the individual paper scores, writing test scores tend to have no or low correlations with academic success . For example, Cotton and Conrow, (1998) found no significant correlation between the participants' IELTS writing scores and their academic achievement. Kerstjen and Nery (2000) reported a correlation of **0.25** between their participants' IELTS writing test and academic scores

▶ Results of this study:

|		Mean real-life score			Mean real-life score
Test Task A total scores (n=160)	Pearson Correlation	.306**	Test Task B total scores (n=140)	Pearson Correlation	.379**
	Sig. (2-tailed)	.000		Sig. (2-tailed)	.000

▶ 32

Findings of criterion-related validity –
Correlation between Test Task A scores and academic
outcome

- ‣ The reading-into-writing test scores seem to be able to 'predict' performance in the target context better at high (Grade B) and low (Grade D) levels than at the mid level.

- ‣ However, for participants who achieved at the mid level (Grade C), their scores on the reading-into-writing test tasks ranged widely. It appears that academic writing ability might have limited impact at the mid-level academic achievement in the context of this study.
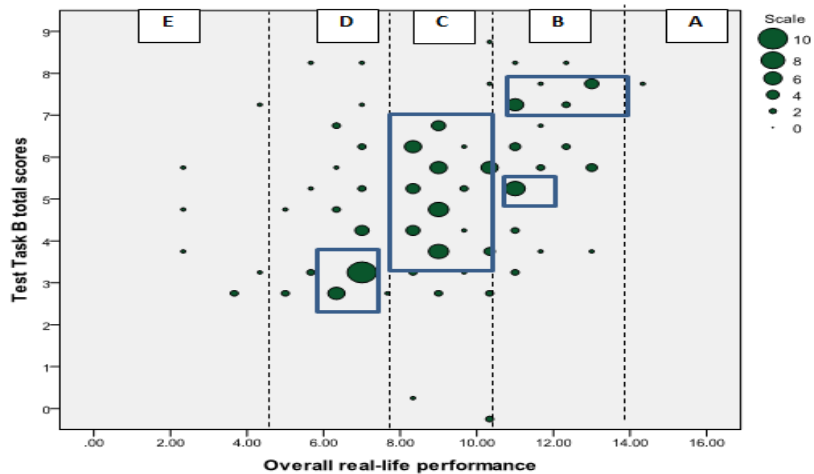
‣ 33

Findings of criterion-related validity –
Pattern of the correlation between Test Task A scores
and academic outcome
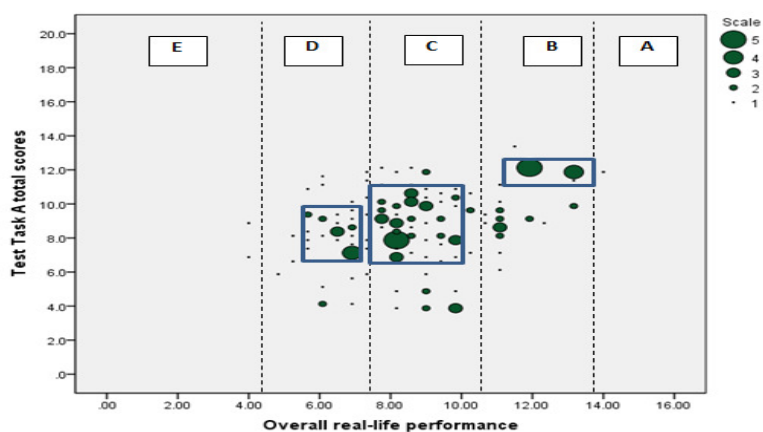
‣ 34

Findings of criterion-related validity –
Pattern of the correlation between Test Task B scores
and academic outcome

35

Implications for test validation

- This study went beyond the scope of the earlier studies to put forwards a framework with explicit **contextual** and **cognitive** parameters for test developers and further researchers .
- The results of this study strongly suggest that the integrated reading-into-writing task type is a valid tool to assess academic writing ability in terms of the context validity, cognitive validity and criterion-related validity.

36

Implications for the development of valid academic writing tests

▶ **Overall task setting**

- Incorporating other common academic writing genres, such as report.
- Avoiding the use of topics in the social domain.
- Incorporating more language functions

▶ **Input text features**

- Incorporating more input genres and a combination of argumentative texts and expository texts.
- Reducing the lexical complexity

▶ 37

Implications for the development of valid academic writing tests

▶ Target cognitive process U2

- This study identified eleven cognitive processes elicited by the real-life academic writing tasks:

  (1) task representation and macro-planning, (2) revising macro plan, (3) connecting and generating, (4) selecting relevant ideas, (5) careful global reading, (6) organising ideas in relation to input texts, (7) organising ideas in relations to own text, (8) low-level editing while writing, (9) low-level editing after writing, (10) high-level editing while writing, and (11) high-level editing after writing

- The results show some discrepancy in the underlying structure of the processes of (1), (5), (7) and (8) between the test and real-life conditions.

▶ 38

U2   A major threat to the cognitive validity of independent writing-only tests is that the tasks might tap into a skill which is solely used under test conditions and demonstrate little relation to the real-life processes.
USER, 11/07/2013

Implications for score interpretation

▸ The two reading-into-writing test scores were able to predict performance in the target context better at high and low levels than at the mid-level.

▸ It appears that academic writing ability might have limited impact on the mid-level academic achievement in the context of this study. Therefore, any high-stakes decisions for these mid-level test takers need to be made with extra caution, and supported by other forms of evidence.

▸ 39

Q & A

Thank you for your attention!

Sathena.chan@beds.ac.uk

▸ 40