A "BEFORE-AFTER" VALIDATION STUDY OF AN ENGLISH LANGUAGE SCREENING TOOL

Cathie Elder & Janet von Randow

University of Melbourne & University of Auckland

EALTA Conference
Sitges, Spain
June 2007

Outline of paper

- INTRODUCTION
- RESEARCH CONTEXT
- RESEARCH QUESTIONS
- METHOD AND RESULTS
- SUMMARY OF FINDINGS
- CONCLUSION

Introduction

- Popularity of performance-based language tests over the past three decades has resulted in indirect tests falling somewhat out of favour

- Such tests remain widely used, particularly in low stakes placement contexts, but scant attention has been paid to their validation *(but see Wall et al. 1994, Fulcher 1997, Fox 2005)*

- Best practice in language testing however demands that all testing initiatives be subject to ongoing validation both at the development stage and after the test has been put to use *(Cronbach, 1988, Alderson et al, 1995, Davies & Elder 2004)*

- This paper describes ongoing efforts to validate an indirect screening test used as part of a post-admission diagnostic procedure designed to identify the academic language and literacy needs of a linguistically and culturally diverse population.

Context for the research (1)

- An English medium university with overall intake of circa 6,500 per annum and a large EAL population, including both domestic (immigrant) students and international ("visa") students.
- Reported difficulties with English amongst many EAL students due to
  - limited exposure/experience of English in academic contexts
  - non stringent English entry requirement for both international & domestic students

Context for the research (2)

- In 2000 the university moved to introduce a post-admission “diagnostic” English language assessment to identify students at risk due to limited command of academic English.
- Policy of universal testing (of learners from all language backgrounds) favoured on legal and equity grounds.
- Context demanded an approach which enabled large numbers of students to be tested quickly and economically.

Design of the assessment

TWO TIER PROCEDURE

Tier One

20 minute machine-scoreable/online screening test

Purpose: to exempt linguistically competent majority from further testing

Tier Two

2 hour diagnostic test (DELNA) of English Listening, Reading and Writing

Purpose: to identify language support needs of non-exempt “at risk” students

Tier Two: Diagnostic assessment

- Adapted version of DELA (University of Melbourne) with the following components
  - Listening (30 mins))
  - Reading (45 mins)
  - Writing (30 mins)

- Is the subject of ongoing validation research reported in Elder & Erlam (2001), Elder et al. (2004), Bright & von Randow (2004), Elder et al (2006), Knoch (2007), Read (forthcoming).

Tier Two: Diagnostic assessment

Results for each skill reported on a 6 point scale

Bands 8 & 9: *No support required. Unlikely to experience any difficulties with academic English*

Band 7: *English is satisfactory and no support is required. May nevertheless benefit from further practice in one or other skill area.*

Band 6: *English is mainly satisfactory but would be advised to seek concurrent support in one or more skill areas.*

Band 5: *May be at risk with academic study due to limited English skills. Needs intensive English language support.*

Band 4: *Is likely to be at severe risk of academic failure due to inadequate English. Needs intensive English language support.*

Tier One: Screening

PART A Academic vocabulary (7 minutes)

27 words from Academic Word List (Beglar and Hunt 1998)

Task is to match word to its synonym selected from a number of multiple-choice distractors

Tier one: Screening

SAMPLE VOCABULARY ITEM

You must choose the right word to go with each meaning, as shown below. Fill the circle for the letter that corresponds to the correct word for each meaning.

A	B	C	D	E	F	
○	○	○	○	○	○	procession
○	○	○	○	○	○	sudden attack
○	○	○	○	○	○	dangerous situation

a. pretend
b. bellow
c. parade
d. raid
e. link
f. emergency

Rationale for choice of vocabulary test

- quick and simple measure of vocabulary knowledge *(Beglar & Hunt 1999)*
- passive vocabulary knowledge is a good predictor of active knowledge *(Laufer, Elder, Hill & Congdon 2004)*
- vocabulary correlates strongly with other language skills *(Read 2000)*
- vocabulary (esp. UWL) is a good predictor of academic performance *(Laufer 1991, Hazenberg & Hulstijn 1996, Ellis & Loewen 2001)*
- ease of scoring (by machine or online)

Tier One:Screening

Part B Speed Reading/Cloze elide (10 minutes)

- 73 item test involving speed reading of a passage with a redundant word inserted randomly in each line.
- Task is to identify the redundant words

Note: redundant words for insertion are randomly selected from a text of similar difficulty.

Tier One:Screening

SAMPLE CLOZE ELIDE TASK

In the following passage there is a word in each line that does not belong. Circle that word in each line and then check your answers with the passage beneath. The word that you should have circled will be in bold print.

Some public libraries have developed ways to personal capture the attention print and interest of younger library patrons. One library in the United Kingdom has space on its website for can teenagers to write book and music reviews you of library materials and they are encouraged to participate in the development of library collections from.

Rationale for choice of cloze elide format

- Assesses skimming and scanning which are important in academic contexts (Gibson & Levin 1975, Alderson 2000),
- Assesses "potential comprehension of the test text" *(Davies 1989)*
- Correlates with measures of reading, listening, speaking and writing *(Manning 1987)*
- Speededness (requires rapid processing and hence draws on automated or implicit language knowledge which will discriminate between EAL and ESB
- Ease of administration and scoring (machine scorable or automated on-line scoring)

Research Questions

1. What are the measurement properties of the 2 screening components?
2. How accurate is the screening as a predictor of language need?
3. How useful is the screening procedure in the University of Auckland context?

Measurement properties

Method

- Preliminary trialling of vocabulary and speed-reading with first year undergraduates (N=101)

L1 English (ESB)	24%
L2 English (EAL)	71%
Background unspecified	5%

- Analysis of responses using classical and Rasch test analysis software

Measurement properties

Descriptive statistics for the vocabulary and speed reading components

	N	Minimum	Maximum	Mean	SD	Alpha
Vocabulary	101	6	27	23	4	0.88
Speedreading (cloze elide)	101	1	73	35	19	0.89

Vocabulary and speed-reading scores by language background

		Vocabulary		**Speed Reading**	
	N	Mean	SD	Mean	SD
L1 English (ESB)	24	25.8	1.2	49.1	15.6
L2 English (EAL) *(more than 2 yrs in NZ)*	31	22.2	3.6	27.7	15.1
L2 English (EAL) (*less than 2 yrs in NZ)*	40	18.9	4.9	20.9	12.4

Predictive power

Method

- Trialling of screening AND diagnostic components with larger & more heterogenous sample)

L1 English (ESB)	51%
L2 English (EAL)	49%

- Correlations between screening and diagnostic components (N=190)
- Logistic regressions with minimum bandscore of 7(diagnostic) as criterion to establish optimum cut-score on the screening (N=353)

Predictive power

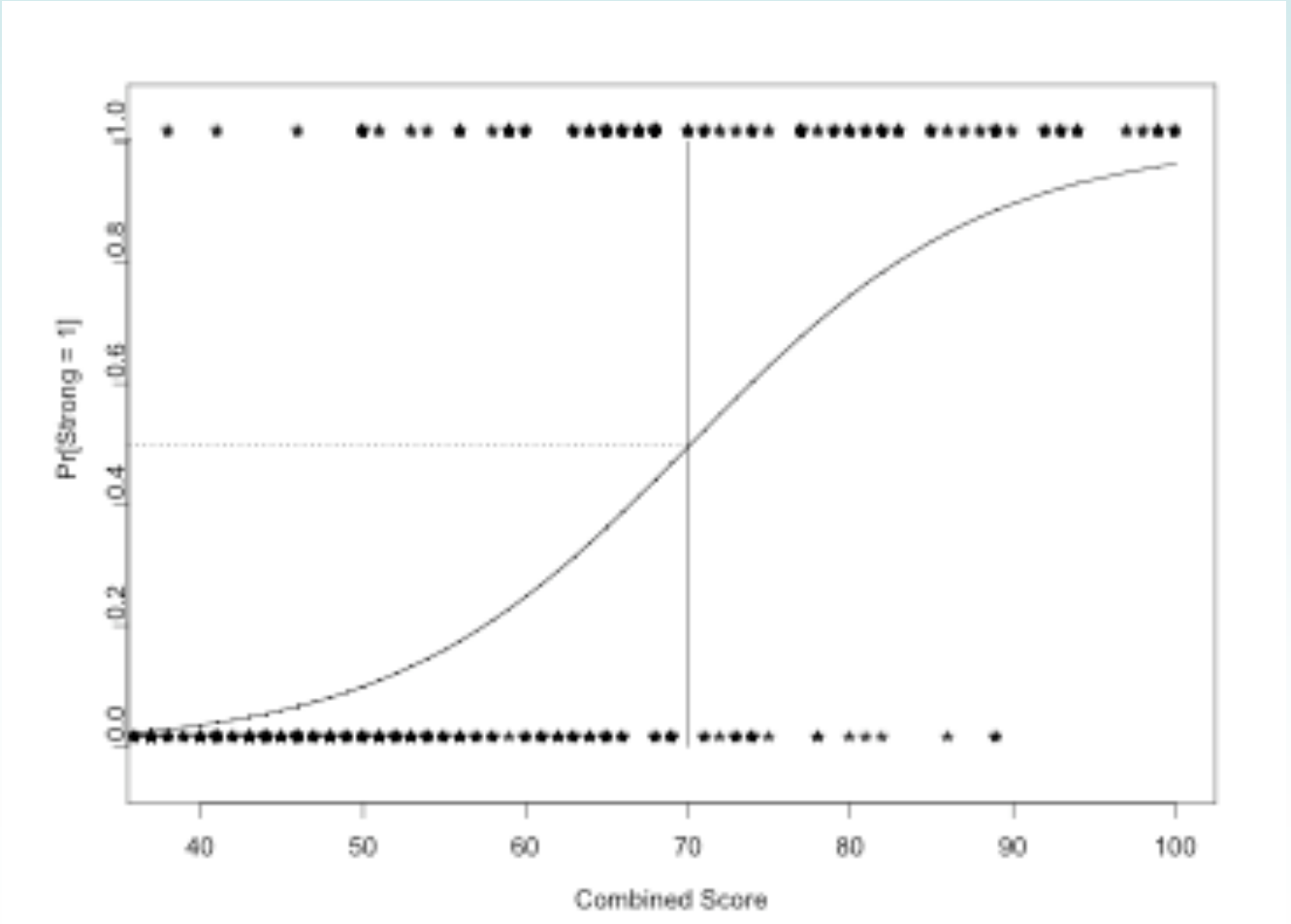
Correlations between screening and diagnostic components

Sub test	**Listening**	**Reading**	**Writing**	**Overall average**
Vocab	.79*	.69*	.56*	.74*
Cloze elide	.71*	.66*	.62*	.77*
Vocab & cloze elide	.76*	.71*	.65*	.82*

Predictive power

Logistic regression

Fitted model



Predictive power

Impact of setting combined vocab and speed reading cutscore at 70 (or above)

	Criterion < Band 7	Criterion >= Band 7	Total
Combined Screening Score < 70	**251**	**37**	**288**
Combined Score >= 70	**19**	**46**	**65**
Total	**270**	**83**	**353**

- Sensitivity (251/270) 93%
- False negatives (37/288) 13%
- Exemption from further diagnosis (65/353) 18%

Predictive power

Impact of setting cutscore at 60 (vocabulary and speed reading combined)

	Criterion < Band 7	Criterion >=Band 7	Total
Combined Score < 60	**218**	**15**	**233**
Combined Score >= 60	**52**	**68**	**120**
Total	**270**	**83**	**353**

- Sensitivity (218/270) 81%
- False negatives (15/233) 0.06%
- Exemption from further diagnosis (120/353) 34%

Cut-score policy

- Combined score of >=70 : exempt from diagnosis
- Score 60-69: also exempt but advised to do independent English study
- Score < 60 strongly recommended to take second-tier diagnosis to obtain profile of language needs and be guided towards appropriate English language support

Utility and relevance

Method

Formulation of validity hypotheses/criteria in relation to test purposes and intended impact *(Davies & Elder, 2004)*

Collection and analysis of operational data

Utility and relevance

Criteria

Uptake: Departments/faculties will adopt the screening

Efficiency: Screening will be useful in exempting substantial numbers from further diagnosis

Equity: Proficient students will be exempted without flagging ethnicity as grounds for special intervention

Return rate: Students identified by screening as “at risk” will return for further diagnosis

Utility and relevance: *uptake*

Uptake and delivery mode of screening and diagnostic assessments : 2002-2006

	2002	2003	2004	2005	2006
Screening & diagnosis ***(in one sitting)***	**231**	**156**			
Screening only ((prior to diagnosis)	**245**	**132**	**2566**	**4046**	**5490**
Total screening	**476**	**288**	**2566**	**4046**	**5490**

Utility and relevance *(efficiency)*

Screening exemption rates

	2002	2003	2004	2005	2006
No taking screening	476	289	2566	4046	5490
Number (%) exempt	227 (48%)	110 (38%)	1639 (64%)	2682 (66%)	4512 (76%)

Relevance and utility (*equity*)

Exemption rate by language background

ESB

	2002	2003	2004	2005	2006
All screening candidates	**223**	**102**	**1109**	**1466**	**3064**
Exempt candidates	**86%**	**65%**	**90%**	**92%**	**93%**
Candidates recalled for diagnosis	**14%**	**35%**	**10%**	**8%**	**7%**

EAL

	2002	2003	2004	2005	2006
All screening candidates	**254**	**187**	**1418**	**1814**	**2428**
Exempt candidates	**26%**	**56%**	**40%**	**46%**	**55%**
Candidates recalled for diagnosis	**74%**	**44%**	**60%**	**54%**	**45%**

Relevance and utility *(return rate)*

Return rates for screening

	2002	2003	2004	2005	2006
Number recalled for diagnosis	**248**	**179**	**927**	**1364**	**1338**
Number (%) return rate	**129 (52%)**	**56 (31%)**	**440 (47%)**	**526 (39%)**	**444 (33%)**

Summary of findings

1. What are the measurement properties of the 2 screening components?

Both vocabulary and speed reading are acceptably reliable measures measures which discriminate between learners with different degrees of exposure to English

Summary of findings

2. **How accurate is the screening as a predictor of language need**?

   In combination they are capable of yielding acceptably accurate predictions of performance level on a diagnostic test of listening, reading and writing for academic purposes.

   Screening cutoff of 70 captures vast majority of linguistically at risk students. The lower cutoff of 60 is less sensitive, but defensible for a low stakes assessment.

Summary of findings

3. **How useful is the screening procedure in the University of Auckland context?**

Uptake: widespread (if not yet universal) use of the screening tool for first year undergraduate intake

Efficiency: high exemption rate results in substantial cost savings to the university.

Summary of findings

Equity

Test has succeeded in its purpose of exempting highly proficient (mainly ESB students) without resorting to discriminatory practice of targeting particular ethnicities for testing.

The screening does not guarantee “immunity” for L1 English (ESB) students, although, as expected there is a much higher % of these than L2 (EAL) speakers among the exempt students.

Summary of findings

Return rate

Poor rate of return of students identified as needing diagnosis seriously undermines the efficacy of the DELNA program in supporting linguistically at risk students

BUT

- Steps taken to address this issue
- Indirect spin-offs from using screening
  - greater awareness of English language issues among staff
  - increased action to address academic literacy needs

Further validation efforts planned or in train

- further attention to test presentation and delivery issues to improve return rate and ensure appropriate interpretations and actions based on screening test scores
- exploration of alternatives to AWL levels test to refine predictive power of the screening
- feedback from faculties to ascertain whether exempt students are linguistically competent
- resetting of cut-scores on screening as new versions are developed and as modifications are made to the follow-up diagnostic procedure