The 23rd Language Testing Symposium ACROLT 2007

A six-point cut-off scale for an online English placement test comprising groups of progressively higher-level test items

Dr Chris Alexander
Dr Papadima-Sophocleous

INTERCOLLEGE

Nicosia

Overview

- A brief description of the test background and stages of test development will precede a detailed discussion of the method for placing students and a description of the nature of NEPTON data.

- In this paper it will be argued that The New English Placement Test Online (NEPTON) places students fairly as it considers the entire performance of a student on all test-item levels. The NEPTON does not target test items at a derived level of the candidate. ECOLT (2006) Tucker and Shark maintain that some researchers publicly 'declare CAT is dead'

- A full paper version of this presentation is available (Alexander and Papadima-Sophocleous 2007).

NEPTON has been informed by language level frameworks such as CEFR (Council of Europe, 2001) and Association of Language testers in Europe (ALTE, 2002). It places students into six levels from beginner to upper-advanced

Intercollege course code	Level	Equivalent CEFR
BENG50 NEPTON level 1	Beginner	A1
BENG80 NEPTON level 2	Elementary	A2
BENG90 NEPTON level 3	Intermediate	B1
BENG100 NEPTON level 4	Upper-intermediate	B2
ENGL100 NEPTON level 5	Advanced	C1
ENGL101 NEPTON level 6	Upper-advanced	C2

Table 1   CEFR equivalents of Intercollege non-elective ESOL courses

A test *question* comprises 1 or 4 to 5 test *items.*
Each *item* consists of mainly 5 multiple-choice **options.**

Each NEPTON test has **33 _questions_** for the test-taker to take, however the test-taker actually responds to **54 test _items_**

http://nepton.engine.intercol.edu – NEPTON Engine – Microsoft Internet Explorer

Questions: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

Letter to a Friend

Read the passage about Letter to a Friend and choose the correct answer from the dropdown menu selection.

Dear Sean,

It was good to see you in Manchester last week. Unfortunately I had a lot of work every day, so I didn't have much time to see you more. I didn't have the time to visit many of the interesting places [▼] are famous in Manchester either. However, I [▼] many useful contacts so I hope my company can do a lot of business in the United Kingdom in the future. I know you don't have much free time during your visits in Cyprus, but please try to come and see us next time you are here. There are a lot of interesting places in Cyprus, too. Do you have many things to do during your next visit? You [▼] come from the beginning of summer because the weather is terrific and when you [▼] with your work we can go to the beach. My [▼] organises bush walking and he can take us to some beautiful places.

wife brother
wife's brother
wives brother
wifes brother
wife brothers

Looking forward to seeing you then,

Love,
Petros.

papas, salomi   Remaining Time: 77 min.   [>]   **Finish Test**

Done   Internet

Plate 1   Test area screenshot

Every NEPTON test has thirty-three questions that are grouped into six, nine-item slides that represent the levels of the six Intercollege ESOL courses described in Table 1.

The abbreviations used in Table 2 are for the NEPTON question types, these types are: sentence-based structure (SB-S), text-based structure (TB-S), sentence-based vocabulary (SB-V), text-based vocabulary (TB-V), sign-based reading comprehension (SB-RC), and text-based reading comprehension (TB-RC). These question types were mainly based on Heaton (1995), Hughes (2000) and Harmer (2001).

	TEST 1	No of items per question type	TEST 2	No of items per question type
NEPTON level 1 **Beginner** 9 items per slide	SB-S: TB-S: SB-RC:	3 5 1	SB-V: TB-V or TB-RC: SB-RC:	4 4 1
NEPTON level 2 **Elementary** 9 items per slide	SB-V: TB-V / TB-RC: SB-RC:	4 4 1	SB-S: TB-S: SB-RC:	3 5 1
NEPTON level 3 **Intermediate** 9 items per slide	SB-S: TB-S:	4 5	SB-V: TB-V or TB-RC:	5 4
NEPTON level 4 **Upper-intermediate** 9 items per slide	SB-V: TB-V or TB-RC:	5 4	SB-S: TB-S:	4 5
NEPTON level 5 **Advanced** 9 items per slide	SB-S: TB-S:	4 5	SB-V: TB-V or TB-RC:	5 4
NEPTON level 6 **upper-advanced** 9 items per slide	SB-V: TB-V or TB-RC:	5 4	SB-S: TB-S:	4 5
9 items x 6 slides	**54 items (33 questions)**		54 items (33 questions)	

Table 2 **The NEPTON item slides**

NEPTON Item writing

- The item topics were similar to topics

- **(1) covered in textbooks of the same levels;**

- **(2) suggested by theoretical books for those levels (Nunan, 1989, 1991; Heaton, 1995; Hughes, 2000; Harmer, 2001);**

- **(3) drawn from the suggested topics of L2 level frameworks such as CEFR, and ALTE;**

- **(4) related to educational settings of students**.

- A variety of format such as sentence, signs and text, and reflecting the respective levels, skills and content are used in the NEPTON test.
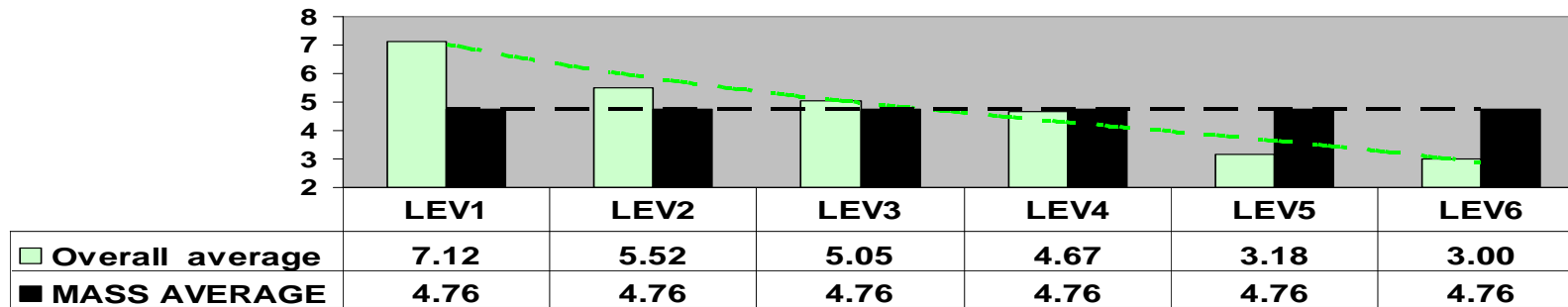
Reliability and Validity

- A large and revised item pool of 1084 items

- Short and clear instructions for the NEPTON test.

- A moderation process took place in the initial stages of the item bank development. This resulted in (1) the elimination of ambiguous items; (2) the agreement of acceptable responses; (3) the editing of questions, texts and multiple-choice options.

- Continuous item analysis is periodically carried out

- External validity considered

- Six full-timers with expertise in the area of testing were also asked to examine NEPTON's content validity

- There was positive correlation with a separately marked hand-written writing component and the NEPTON (Weir, 1990; (Heaton, 1995; Hughes, 2000; Alderson et al. 2001; Bachman, 2003; Alderson and Bachman 2004).

- Even though test-takers have a choice of taking the test electronically or in pen-and-paper format very few take the NEPTON pen-and-paper option. This finding indicates that test-takers are familiar with computers and have positive attitudes towards the computer delivery of the NEPTON; this finding also suggests that the test has strong internal face validity

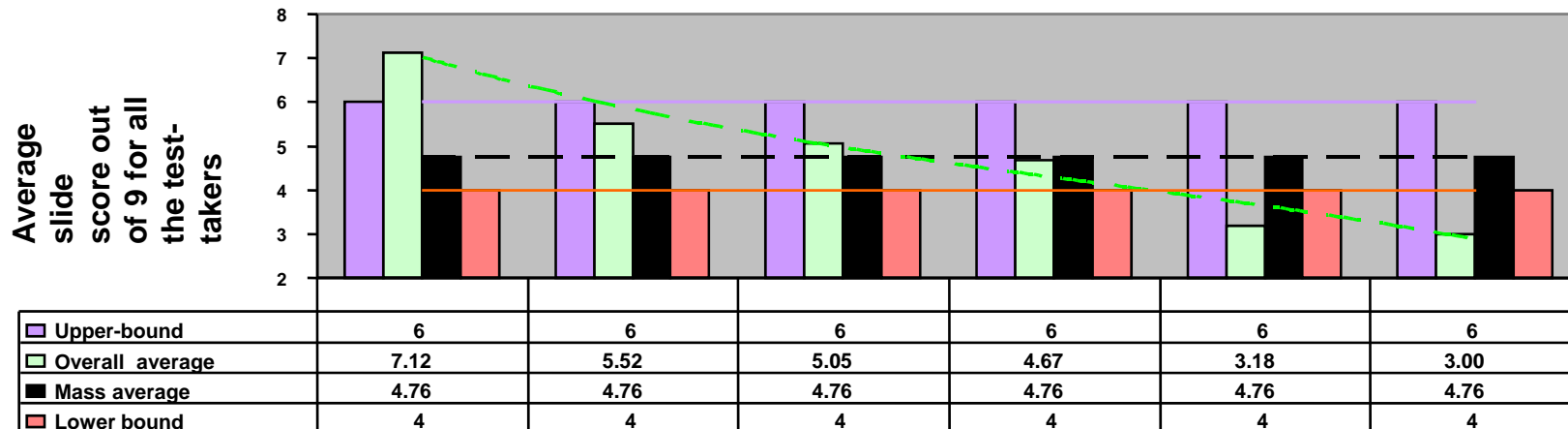- The pre-test e-tutorial; available on http://nepton.engine.intercol.edu

Calculating the slide cut-offs

Six overall performance averages of all testees on all NEPTON levels and the mass average of these six averages
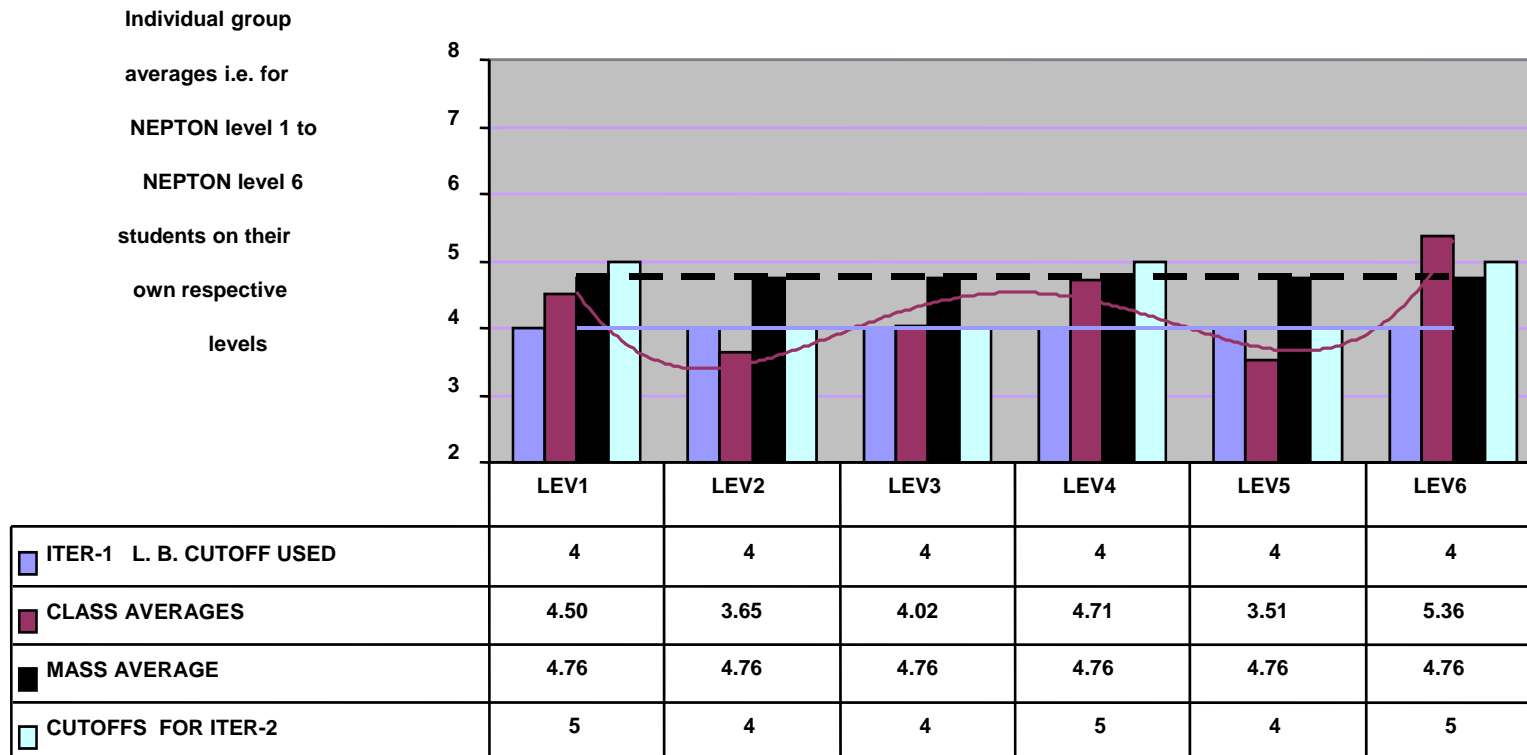

Averages out of nine

	LEV1	LEV2	LEV3	LEV4	LEV5	LEV6
▫ Overall average	7.12	5.52	5.05	4.67	3.18	3.00
■ MASS AVERAGE	4.76	4.76	4.76	4.76	4.76	4.76

Upper (6) and lower (4) bounds set above and below the overall integer slide average of 5

Average slide score out of 9 for all the test-takers

▫ Upper-bound	6	6	6	6	6	6
▫ Overall average	7.12	5.52	5.05	4.67	3.18	3.00
■ Mass average	4.76	4.76	4.76	4.76	4.76	4.76
▫ Lower bound	4	4	4	4	4	4

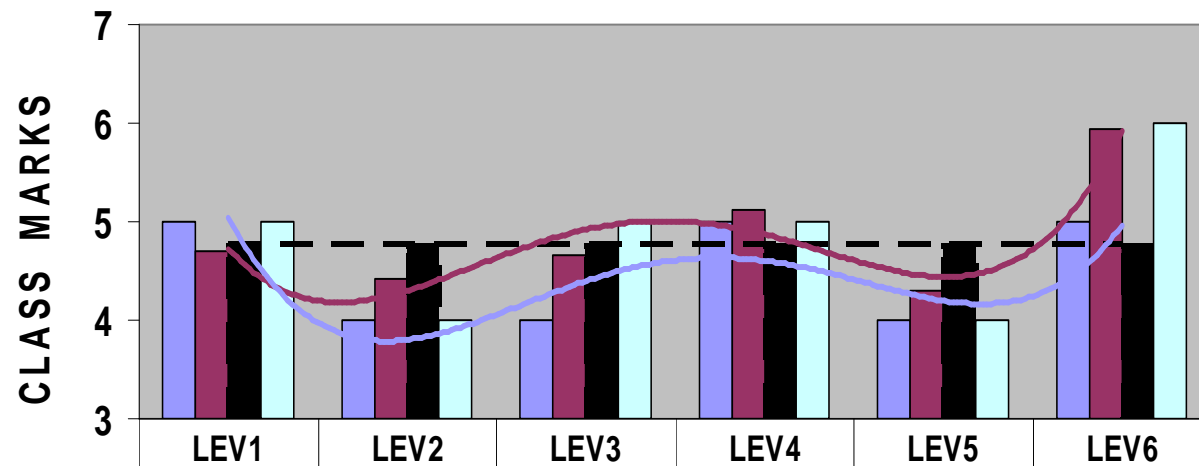A second iteration using the new first iteration **5-4-4-5-4-5** slide cut-offs led to new integer slide cut-offs of **5-4-5-5-4-6**. These new second iteration cut-offs were used in a further third iteration. The results were **5-4-5-5-6-7**. This third iteration led to an unacceptably high ENGL101 (level 6) cut-off of seven, which went above the upper-bound average condition of 6 discussed above and set at the beginning of the iteration process. Therefore the second iteration cut-offs of **5-4-5-5-4-6** had to be rejected and the new NEPTON cut-offs were therefore the first iteration cut-offs of **5-4-4-5-4-5.**

Lower bound cut-off (4) used in the first iteration to calculate new iterated slide cut-off points

Individual group averages i.e. for NEPTON level 1 to NEPTON level 6 students on their own respective levels

	LEV1	LEV2	LEV3	LEV4	LEV5	LEV6
ITER-1 L. B. CUTOFF USED	4	4	4	4	4	4
CLASS AVERAGES	4.50	3.65	4.02	4.71	3.51	5.36
MASS AVERAGE	4.76	4.76	4.76	4.76	4.76	4.76
CUTOFFS FOR ITER-2	5	4	4	5	4	5

Second iteration

NEW Cut-offs of 5-4-4-5-4-5 used in second iteration

CLASS MARKS

	LEV1	LEV2	LEV3	LEV4	LEV5	LEV6
ITER-2 CUTOFFS FROM ITER-1	5	4	4	5	4	5
CLASS AVERAGES	4.70	4.43	4.66	5.13	4.31	5.94
MASS AVERAGE	4.76	4.76	4.76	4.76	4.76	4.76
FOUND CUTOFFS FOR ITER-3	5	4	5	5	4	6

Third iteration

New cut-offs of 5-4-5-5-4-6 used in third iteration

NOT PERMISSIBLE AS IT EXCEEDS UPPER BOUND HENCE USE CUTOFFS OF ITER-2 AND STOP
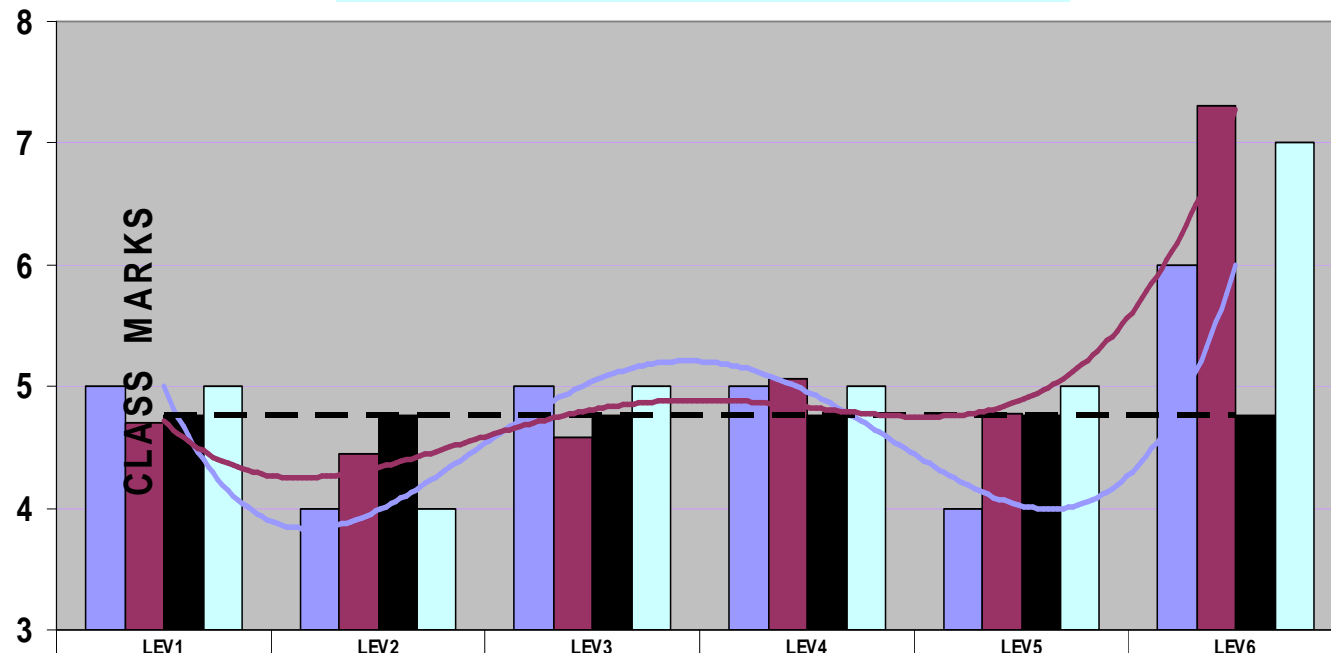
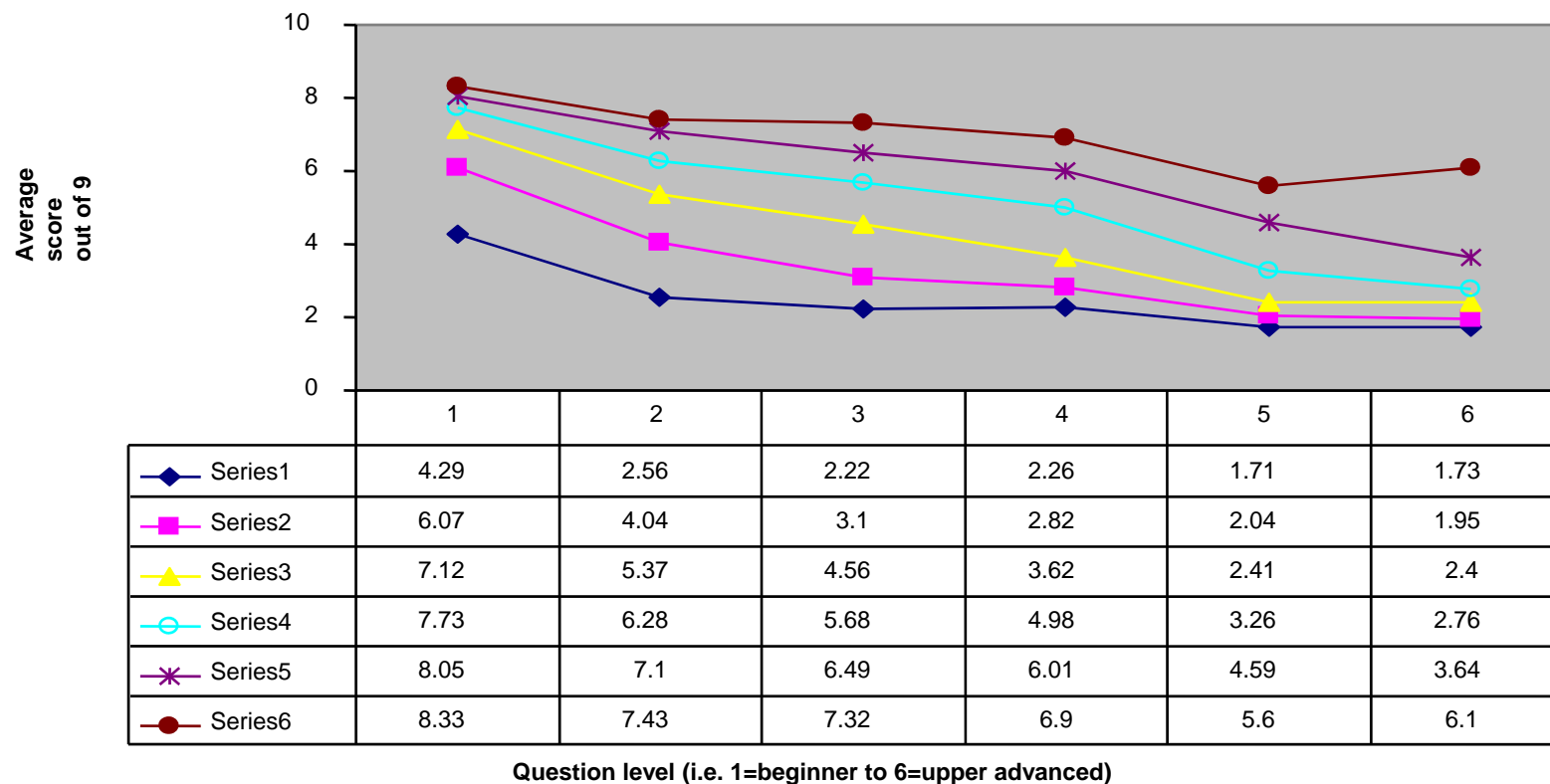|	LEV1	LEV2	LEV3	LEV4	LEV5	LEV6
ITER-3 CUTOFFS FROM ITER-2	5	4	5	5	4	6
CLASS AVERAGES	4.70	4.45	4.58	5.06	4.77	7.3
MASS AVERAGE	4.76	4.76	4.76	4.76	4.76	4.76
FOUND CUTOFFS FOR ITER-4	5	4	5	5	5	7

Each slide therefore has a cut off point. If the cut-off point per slide is achieved, the student is deemed to have fulfilled the placement criterion of that individual slide, and is awarded one point; there are 6 points in total. If the student scores 0-1 points in total on any slide in any order the placement is NEPTON level 1, if the student scores 2 points in total on any two slides in any order the placement is NEPTON level 2. This continues to a maximum of six points.

	NEPTON level 1	NEPTON level 2	NEPTON level 3	NEPTON level 4	NEPTON level 5	NEPTON level 6	Total number of points	Final NEPTON level
NEPTON Cut-offs	<u>5</u>	<u>4</u>	<u>4</u>	<u>5</u>	<u>4</u>	<u>5</u>		
	3	3	2	2	0	1	0	1
	8	3	0	0	0	2	1	1
	7	9	3	3	3	1	2	2
	9	2	8	6	2	2	3	3
	8	6	8	9	2	0	4	4
	8	9	7	5	8	1	5	5
	9	8	9	9	9	9	6	6

Table 3   Examples of student slide performance and their corresponding placement using 5-4- 4-5-4-5 slide cut-offs

The 2005-6 NEPTON data analysis of 2063 students

Performance of all NEPTON level 1 students to NEPTON

level 6 students on all the questions (i.e. from beginner

to upper-advanced level questions

	1	2	3	4	5	6
Series1	4.29	2.56	2.22	2.26	1.71	1.73
Series2	6.07	4.04	3.1	2.82	2.04	1.95
Series3	7.12	5.37	4.56	3.62	2.41	2.4
Series4	7.73	6.28	5.68	4.98	3.26	2.76
Series5	8.05	7.1	6.49	6.01	4.59	3.64
Series6	8.33	7.43	7.32	6.9	5.6	6.1

Average score out of 9

Question level (i.e. 1=beginner to 6=upper advanced)

The 2005-6 NEPTON data analysis of 2063 students

- Higher-level student slide averages are always better than lower-level student slide averages. Test items therefore discriminate well globally.

- Series 1 (NEPTON level 1) students on slides three to six on average score 2.22, 2.26, 1.71, 1.73 respectively. This oscillates around two, which is the approximate slide guess factor of twenty percent (2 out 9 is approximately 20%).

- Series 6, 5, and 4 student averages drop slowly on low-level slides;

- Overall slide averages presented above suggest that slide item 'difficulty' is 'relative' and depends on the level of the test-taker. with the exception of series-six students, that slide-item difficulty on average is noticeable on the student's actual NEPTON slide level or on higher slide levels.
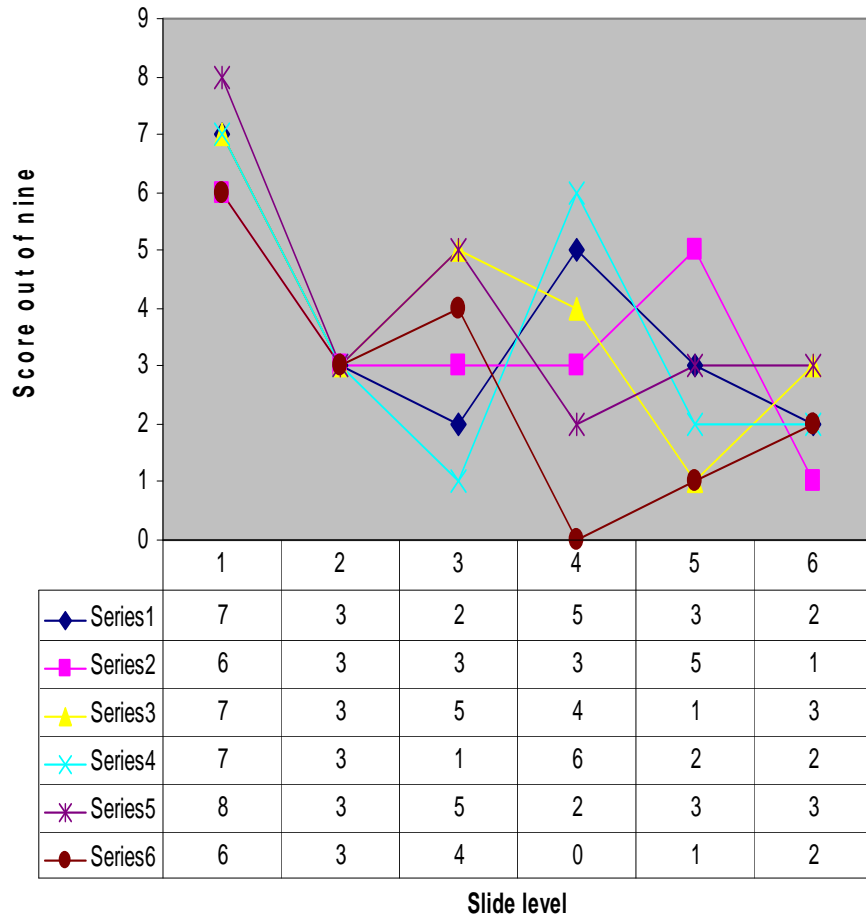
Even though the NEPTON overall averages follow the acceptable trends described in section VII, individual student slide performance on the NEPTON is usually erratic. Students individually often score higher on higher-level slides despite regular item analyses and test-item moderation.

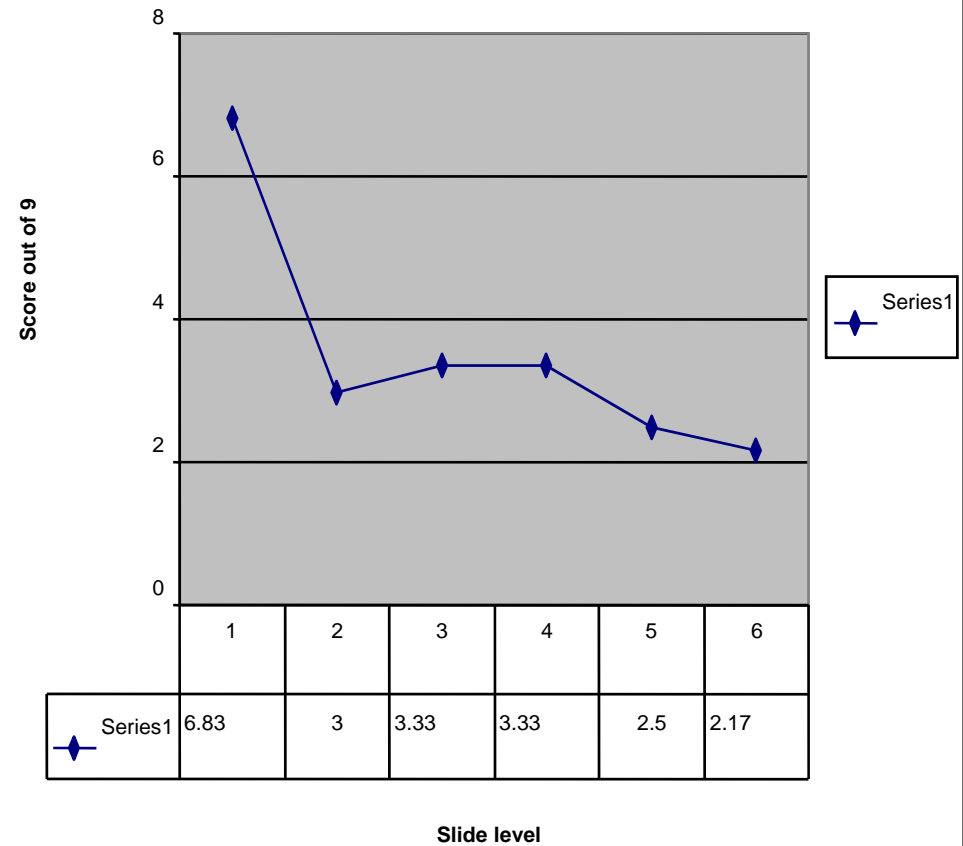| NEPTON STUDENT LEVEL	Number of students at this NEPTON level that scored higher on higher-level slides at some point during the test	Number of students in the database at this level	Percentage of students that scored higher
1	221	243	91%
2	224	234	96%
3	303	370	82%
4	433	484	90%
5	353	426	83%
6	276	306	90%

Table 4   Students that scored higher on higher-levels NEPTON slides for 2005-6

A sample of NEPTON level-2 student slide performance. It indicates visually how erratic student slide performance. However the figure on the right plots the averages of the six erratic sample student performances. Bearing in mind this sample size, the averages are close to the overall level-2 averages presented in Figure 5 (6.07, 4.04, 3.1, 2.82, 2.04, 1.95).

6 Examples of NEPTON level 2 student slide performance i.e. slide performance in the NEPTON is very often erratic


	1	2	3	4	5	6
Series1	7	3	2	5	3	2
Series2	6	3	3	3	5	1
Series3	7	3	5	4	1	3
Series4	7	3	1	6	2	2
Series5	8	3	5	2	3	3
Series6	6	3	4	0	1	2

Slide level

Average performance of the six 'erratic' examples in Figure 6

	1	2	3	4	5	6
Series1	6.83	3	3.33	3.33	2.5	2.17

Slide level

LEVEL 2 Globally

Performance of elementary students on Beginner to Upper-Advanced questions

Average score out of 9 for 234 elementary students

	1	2	3	4	5	6
Series1	6.06838	4.04274	3.09829	2.81624	2.03846	1.95
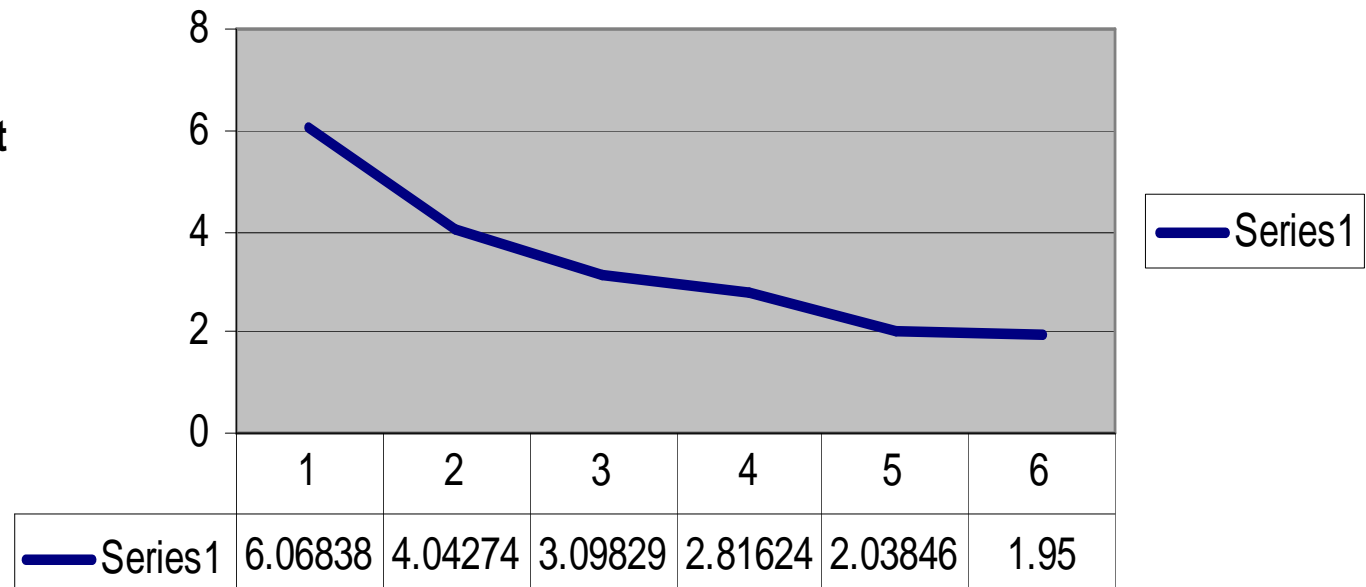
Level of question i.e. 1 easiest to 6 hardest

A more striking aspect of the data is how approximately between twelve and nineteen percent of NEPTON level 1 to NEPTON level 4 students can achieve slide cut-offs on items that are two to four levels higher than their NEPTON placement level.

NEPTON level of student	Number of students at this level that achieve the cut-off two levels higher	Number of students at this level that achieve the cut-off three levels higher	Number of students at this level that achieve the cut-off four levels higher	Total integer percentage of students at this level that achieved cut-offs on slides two or more levels higher
1	19 (NEPTON level 3 items)	2 (NEPTON level 4 items)	9 (NEPTON level 5 items)	13%
2	15 (NEPTON level 4 items)	23 (NEPTON level 5 items)	2 (NEPTON level 6 items)	17%
3	51 (NEPTON level 5 items)	18 (NEPTON level 6 items)	NA	19%
4	57 (NEPTON level 6 items)	NA	NA	12%

Table 5   Number of students that obtain cut-offs on slides that are two or more levels higher than their NEPTON level

The six examples in Table 6 are data taken from the 2005-2006 data analysis; they indicate how students can perform significantly better on questions that are two to four levels above their NEPTON level. Students in the NEPTON are expected to answer all the test questions as their complete test performance on all test items is considered by the NEPTON; the NEPTON items therefore are not targeted at the level of the candidate as for instance in Maycock (2007, 7).

	NEPTON Level 1	NEPTON Level 2	NEPTON Level 3	NEPTON Level 4	NEPTON Level 5	NEPTON Level 6	Total number of slide points	Final NEPTON level
NEPTON Cut-offs	**5**	**4**	**4**	**5**	**4**	**5**		
I	3	2	1	3	**6**	2	1	1
II	7	1	2	3	1	**6**	2	2
III	6	5	1	2	2	**6**	3	3
IV	9	8	5	4	3	**8**	4	4
V	8	7	3	7	4	**8**	5	5
VI	9	4	8	7	4	**9**	6	6

Table 6   Six examples of unusual student slide performance

Key Conclusions

- Students in the NEPTON as exemplified in Table 6 can achieve unusual slide cut-offs.

- The decision therefore to award one point per slide even though higher level slides contained higher-level items is a key aspect of the NEPTON slide cut-off system.

- It is held that this system is a fair and effective system as it takes into account the fact that

- **(a) difficulty level averages depend on the NEPTON level of the test-takers;**

- **(b) lower-level students can perform better on higher level questions and that such performance no matter how unusual should, be considered. Moreover, this finding resonates with Ellis's (1985, 118) view of how the dynamic quality of interlanguage is reflected in 'the tremendous variability in learner language and also in overlapping stages of development';**

- **(c) such erratic performance does not skew overall slide averages**

Future research

- More research however is required to determine why individual student slide performance in the NEPTON can be so erratic;

- the following hypotheses are put forward: a combination of student variable competence and/or guessing could be the reason why some students perform significantly better on higher-level questions.

- It is however unlikely that good guessing alone is the reason for the overwhelming percentages of students performing better on higher-level questions as the overall averages would have been skewed.

How are languages learned?
How should language be measured in a placement test?

- Do some learners possess knowledge of advanced grammars and vocabularies? Is SLA an erratic process?

- Should a placement test *attempt to* measure a learner's interlanguage at a **derived** level only (i.e. adaptively)?

- Should a placement test consider the entire performance of a student at all levels?

References

- Alderson, J.C., Clapham, C. and Wall, D. 2001: Language test construction and evaluation. Cambridge: Cambridge University Press.

- Bachman L. F. 2004: Statistical Analyses for Language Assessment. Cambridge University Press.

- Ellis, R. 1985: Understanding Second Language Acquisition. Oxford: Oxford University Press

- Harmer J 2001: The practice of English language teaching. Longman.

- Heaton, J.B. 1995: Writing English language tests. UK: Longman.

- Hughes, A. 2000: Testing for language teachers. Cambridge: Cambridge University Press.

- Maycock, L. 2007: Using simulation to inform item bank construction for the BULATS computer adaptive test. University of Cambridge ESOL Examinations Research Notes, Issue 27, February 2007.Cambridge: University of Cambridge Local Examinations Syndicate and Cambridge University Press.

- Nunan, D. 1991: Language teaching methodology. Hertfordshire: Phoenix ELT

- Weir, C.J. 1995: Understanding and developing language tests. Great Britain: Phoenix ELT.

The NEPTON team

- The NEPTON is the first online English Placement Test developed in Cyprus. It was commissioned by Intercollege (http://www.intercollege.ac.cy), a leading private tertiary institution in Cyprus in 2003.

- The project leader was Dr Papadima-Sophocleous who worked mainly on test content, key features of test interface and areas of item analysis.

- Dr Alexander mainly designed a graded-difficulty test-slide cut-off system and analysed key data derived from this test-slide system. Professor Andreas Alexandrou assisted with the iteration process.

- The programming was undertaken by Dr Dmitry Apraksin. The test updated the existing pen-and-paper English placement practices at Intercollege and brought about change and improvement in test content, delivery and administration.

The 2005-6 NEPTON data breakdown for 2063 students

NEPTON level placement	Number of students in database
Level 1 (BENG50)	243
Level 2 (BENG80)	234
Level 3 (BENG90)	370
Level 4 (BENG100)	484
Level 5 (ENGL100)	426
Level 6 (ENGL101)	306