

DEVELOPMENT AND VALIDATION OF THE DIAGNOSTIC ENGLISH LANGUAGE NEEDS ASSESSMENT (DELNA)

FINAL REPORT

(CONFIDENTIAL)

December 2001

Prepared by

Catherine Elder and Rosemary Erlam
Department of Applied Language Studies and Linguistics
University of Auckland

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	1
2	BACKGROUND	3
3	THE DESIGN AND CONTENT OF DELNA	6
3.1	Screening components	7
3.1.1	Text editing	7
3.1.2	Vocabulary	7
3.1.3	Report writing	8
3.2	Diagnostic components.....	8
3.2.1	Reading comprehension	8
3.2.2	*Listening comprehension	8
3.2.3	*Argumentative essay	9
4	THE AUCKLAND PILOT	10
4.1	Trial of the new assessment components.....	10
4.1.1	Item analysis.....	10
4.1.2	Participant feedback	11
4.2	Trial of the entire assessment battery	12
4.2.1	Participants.....	12
4.2.2	Administration procedures	13
4.2.3	Rating procedures.....	13
4.2.4	Quality of the assessment tasks	13
4.2.5	Performance of the trial candidature	16
4.2.6	Profile of English language needs	18
4.2.7	Initial proficiency screening.....	21
4.2.8	Relationship between DELNA assessment results and University grades.....	22
4.2.9	Reporting and uptake.....	23
5	DELNA DELIVERY	25
5.1	Proposed exemptions	25
5.2	Feasibility	26
5.2.1	Identification of the target candidature.....	26
5.2.2	Timing	27
5.2.3	Reporting.....	27
5.2.4	English language support	27
6	RECOMMENDATIONS	29
7	REFERENCES.....	31
	APPENDIX 1: DIAGNOSTIC PROFILES OF WRITING PERFORMANCE.....	33
	APPENDIX 2: BACKGROUND INFORMATION QUESTIONNAIRE.....	34
	APPENDIX 3: TEST-TAKER FEEDBACK QUESTIONNAIRE.....	35

1 EXECUTIVE SUMMARY

This report describes the process of developing and trialling a Diagnostic English Language Needs Assessment (DELNA) to measure the English language ability of incoming students at the University of Auckland. It should be emphasized that DELNA is intended not as a selection tool but rather as a diagnostic procedure for early identification of students who may be at risk in their studies due to limited English ability.

The procedure, which is an expanded version of a similar instrument already in use at another *Universitas 21* institution, the University of Melbourne, has been piloted on a sample of over 200 undergraduates at the University of Auckland. The sample included substantial numbers of students from both English-speaking and non-English-speaking backgrounds, drawn mainly from populations reported to be experiencing difficulties with English. Results revealed that around two-thirds of those assessed would benefit from further English language support in one or more skill area. The non-English-speaking background students (hereafter NESB) performed at much lower levels than those whose first language was English (hereafter ESB), and many appear to be ill-equipped to cope with the language demands of academia. Some of the ESB students also showed significant weaknesses in academic English, particularly as far as reading ability was concerned. Their needs were, however, far less pressing than those of the NESB students. In general, the DELNA was found to be an effective means of identifying "at risk" students. There were, however, some problems with providing valid diagnostic information for the two groups simultaneously. For example, the characteristics of writing produced by a native and non-native speaker of English with the same overall writing score, were likely to be somewhat different, with, for example, the latter more likely to contain grammatical inaccuracies than the former. Separate scales may therefore need to be developed for reporting purposes and different kinds of language support provided for ESB and NESB students.

It is worth noting that the relationship between scores achieved on DELNA and students' academic performance in various subject areas was stronger than that recorded in most other predictive validation studies, perhaps because of the unusually wide range of language abilities within the trial student population. This may stem from the fact that there is currently no English language entry requirement at the University, except in relation to International students. Further, while University policy officially requires International students to achieve a specified threshold on the IELTS or TOEFL, this policy is unable to be consistently implemented because of difficulties in collecting the necessary data and because of disparities with respect to requirements across different Faculties.

To accommodate this very wide range of language abilities and to streamline the process of assessing large numbers of students within a limited time span, DELNA has been designed as a two-part procedure. The first part serves to screen out students whose performance is clearly adequate. The second part serves to diagnose language strengths and weaknesses for students falling at or below a

certain threshold of ability on Part 1. These students will be issued with a profile of their reading, listening and writing ability, together with information about avenues for English language support.

Recommendation 1

The task of delivering DELNA to a large population of students requires urgent attention and efficient management in order for such an operation to be practicable. It is therefore recommended that an administrator be appointed to liaise with Faculty representatives and to coordinate DELNA delivery and marking, reporting and on-going maintenance.

Recommendation 2

Furthermore, there is an urgent need for concerted university-wide action to set up short- and long-term strategies for delivering appropriate English language support including courses which are tailored to meet the needs identified by DELNA, together with mechanisms for maximising student uptake. Current undergraduate courses in English language only go part of the way towards meeting these needs. They are also inaccessible to many of the students who might benefit from them because some degree programmes make no provision for students to take English/ESOL credit-bearing electives.

Recommendation 3

In the absence of a) a foolproof method of establishing eligibility for DELNA and b) a suitable range of strategies to meet the needs identified by the procedure, it will not be feasible to administer DELNA as a mandatory requirement in the coming year. It is recommended that, in the interim, the needs of the NESB speaker group be prioritized. Available funds should be used in the first instance to administer the procedure to those from non English-speaking backgrounds (according to information supplied on the Admissions form) and who cannot produce acceptable evidence of English proficiency as specified at the end of this report.

Recommendation 4

In addition, special administrations of DELNA should be organized on demand for those academic departments where a) there is particular concern about the English language levels of the students and b) mechanisms for attending to these students' needs have already been identified.

Recommendation 5

Finally, attention needs to be given to improving the quality and consistency of information recorded on *nDeva* regarding the English language qualifications of applicants to the university, so that this information is readily retrievable for the 2003 intake. Without such information, it will be impossible to implement DELNA on a mandatory basis university-wide and, accordingly, to effectively address the English language problems of the kind documented in this report.

2 BACKGROUND

The Diagnostic English Language Needs Assessment (DELNA) is an instrument for measuring students' English language competence. It is designed for incoming undergraduate students at the University of Auckland who cannot show evidence of English language proficiency and may therefore be at risk in their academic studies due to limited English.

For a number of years the University has recognized that limited English may be hampering the academic progress of a substantial number of incoming undergraduates whether or not they are native speakers of English.

In September 1995 the Deans Committee received a report from a Sub-committee on English Language and Entrance, the 'Moran Report' (Moran, 1995). This report recommended, *inter alia*, that

- the University attempt to have seventh form English included in the common University entrance standard;
- that IELTS and TOEFL scores be increased for international students;
- that the University provide funding for credit and non-credit courses in English Language Acquisition.

In 1997 the CUAP Sub-committee on University Entrance discussed English language competence and concluded that under the current legislation Universities could not impose English language criteria on the entry of domestic students.

Following this, the University adopted the practice of asking domestic students from non-English speaking backgrounds who had not received their last two years of schooling in English to self-identify on their application to enroll. Such students were invited to undertake a test in English proficiency and were counseled to take appropriate non-credit or credit remedial courses. This policy had a low compliance rate.

In December 1998, Professor Ellis, Director of the Institute for Language Teaching and Learning (now the Department of Applied Language Studies and Linguistics), proposed that the University apply an English language proficiency entrance examination (Ellis, 1998). This proposal was referred to Faculties, a number of which were unwilling to adopt the proposal.

During 1999 the common University entrance standard became the subject of discussion at a national level in the context of the changing senior secondary school qualification system. Discussion of a literacy standard as part of the common entrance requirement began. This discussion is now drawing to a close and it seems likely that, from 2005, a literacy standard at level 2 of the National Certificate of Educational Achievement will be required for all students entering University with the common

entrance qualification. There will still however be domestic students who enter the University with other qualifications and who may not have achieved this literacy standard.

In December 1999 the Language Teaching and Linguistics Committee of the Faculty of Arts recommended that:

- all entering students who cannot demonstrate satisfactory English proficiency be required to take a Diagnostic English Language Needs Assessment;
- all students be required to take and pass a 'writing intensive course' in their final year at the University;
- a committee be established to identify the kinds and number of English language support programmes needed.

The University accepted the first and third of these recommendations but not the second.

In February 2000 Professor Irene Clark of the University of Southern California reported to the Vice-Chancellor on English language at Auckland University. Professor Clark recommended that a language-based examination administered by the University be used for the purpose of diagnosis and placement rather than for selection, and that both credit and non-credit bearing courses in academic writing be instituted to accommodate the student language needs as identified by the examination.

Accordingly, during 2000, a review of English language entry requirements and English language support mechanisms at 4 other *Universitas 21* institutions was undertaken. The report concluded that the University of Auckland was unique among the *Universitas 21* institutions surveyed and the needs of its students particularly pressing considering that:

- a) there is no English language entry requirement for locally qualified students;
- b) there is an unusually high (compared to other universities) proportion of locally qualified students from non-English speaking backgrounds who, because they have been in New Zealand for more than two years, are NOT required to meet the minimum English standards required for international students;
- c) there are currently no English diagnostic procedures in place for incoming students, either from English or non-English speaking backgrounds;
- d) English enhancement courses are not a mandatory requirement for either undergraduate or postgraduate students.

A proposal for the development and validation of diagnostic assessment procedures was prepared and a grant from the Vice-Chancellor's Development Fund has enabled these procedures to be developed and piloted.

This report gives an account of the content of DELNA (Section 3) and the process of instrument development and validation (Section 4). In Section 5 the practical implications of the pilot process are considered including the feasibility of the proposed exemptions policy, the logistics of DELNA delivery in 2002 and beyond, and, finally, a discussion of possible avenues for English language support linked to the information which DELNA provides.

3 THE DESIGN AND CONTENT OF DELNA

The design and content of DELNA was determined on the basis of a review of the literature (e.g. Weir, 1983; Hughes, 1988; Fulcher, 1997; Rosenfield, Leung & Oltman, 1991; Wall, Clapham & Alderson, 1994). This informed the decision to develop a *general* measure of academic English, rather than producing a series of *discipline-specific* modules (i.e. with texts and tasks tailored to the linguistic demands of particular discipline areas).¹ Care was taken to choose topics which span a range of discipline areas, but which do not require subject-specific background knowledge. A survey of diagnostic English language assessment procedures at other *Universitas 21* institutions confirmed the findings of Gravatt, Richard & Lewis (1997) that the most important language skills required for first year academic study across a range of subject specialisations are reading and listening, followed by writing, with speaking having the lowest priority. Although the importance attributed to each skill differs to some extent between Faculties and course types, it was clearly essential to include measures of reading, writing and listening as part of the assessment. It was also deemed practical to include a discrete-point measure of vocabulary in the battery since vocabulary knowledge is heavily implicated in the effective use of academic English (e.g. see Loewen & Ellis, 2001). Following a model adopted by the Faculty of Medicine at the University of Melbourne, the intention was to establish a two-tier procedure, so that those performing at high levels on the discrete point machine-scorable tasks could be advised that their English was satisfactory and exempted from being assessed on the remaining diagnostic components. Given that the sample was likely to include substantial numbers of students who were highly proficient in English, it was considered that this approach would allow efficient processing of the high volume of students likely to present for the needs assessment by reducing the marking load (and therefore costs) associated with the more labour-intensive listening, reading and writing tasks. One of the purposes of the trial was to establish whether this preliminary screening process was in fact feasible, and whether the measures included in the first part of the procedure were sufficiently accurate in their predictions to allow such exemptions to be made with a reasonably high degree of confidence.

It was also decided, following the review of instruments used at other *Universitas 21* institutions (see above), that some previously validated assessment tasks used for diagnostic purposes at the University of Melbourne (Language Testing Centre 1991) would be suitable for use in the Auckland context. Using the same tasks would have the added advantage of allowing English language standards at the two institutions to be benchmarked against one another. A site licence was therefore purchased to use the listening, reading and writing components of the University Test of English as a Second Language (UTESL), which exists in a number of different versions. The agreement drawn up between the two universities allowed for the use of existing versions and the development of one new version of either the listening or the reading each year, with one party taking responsibility for item writing and the

¹ Research on the predictive power of discipline-specific tests (versus general proficiency tests) has produced equivocal findings (e.g. Clapham 1996). Moreover, equivalence across different modules is difficult to establish and the development, administration and marking of multiple assessment options is labour-intensive.

other for trialling and item analysis. (The assessment components, which are subject to this agreement, are marked below with an asterisk.) It was also agreed that, for the purpose of the Auckland trials, assessments of the subjectively-scored writing scripts would be undertaken, in the first instance, by trained raters at the University of Melbourne so that the cut-offs between satisfactory and unsatisfactory performance and band scores indicating different levels of need could be pegged to standards already set at another *Universitas 21* institution.

An outline of each component of the trial version of DELNA is provided below.

3.1 Screening components

3.1.1 Text editing

This is a 73-item measure which adopts the cloze elide technique (Davies, 1975, 1989; Alderson, 2000) and measures the ability to scan and decode a text under time pressure. A good implicit knowledge of sentence-level grammar and an understanding of English word collocations are an essential prerequisite for successful performance on this component. Students are required to speed read a text that has been “doctored” to include an additional word in each line² and to delete each word “that does not belong”. The text is on a topic of general academic English. Students are awarded a point for every word correctly deleted and no points are deducted for incorrect deletions. Answers can be machine-scored, making the instrument highly efficient to administer.

Only one version of this component has been developed thus far, pending a decision about its utility as a preliminary screening device.

3.1.2 Vocabulary

This is a 27-item measure of receptive knowledge of academic vocabulary (Beglar & Hunt, 1999). Each item contains three definitions, each of which must be matched with the appropriate word chosen from a list of six. Students receive one point for every correct choice. The instrument currently exists in two parallel versions, Form A and Form B. Form A was used for the purpose of the trial (see below).

The decision to use this kind of assessment procedure, rather than another more elaborate and time-consuming vocabulary measure (see Read, 2000 for a review of the range of procedures currently available), was that it had already been trialled at this university and was found to be a robust predictor of performance in academic English for credit courses as well as correlating significantly with students’ overall GPA (Loewen & Ellis, 2001)³. The instrument is also quick and easy to administer

² The additional words are inserted rationally, rather than randomly, with care taken to ensure that the inserted words do not appear at plausible points in the passage.

³ It was decided to use only the University Word List component given that this yielded higher correlations than the 2,000 Word Level component and indeed the two components combined.

and answers can be machine-scored. The test is not however secure and further parallel versions will need to be developed for subsequent administrations.

3.1.3 Report writing

This task requires students to produce a page of commentary based on simple factual information presented in tabular or diagrammatic form. The ability to understand and interpret this kind of information is generally accepted to be a key component of academic literacy.

Scripts are scored holistically on a 4-point scale with 4 representing highly competent performance and 1 representing inability to meet the demands of the task. The scale is designed to allow a quick decision to be made as to the adequacy (or otherwise) of students' writing for academic purposes. The scale in its current form does not however provide diagnostic information about areas of strengths and weakness. It is intended as a means of corroborating information made by the other two screening components as to whether students should proceed to further diagnostic assessment.

3.2 Diagnostic components

3.2.1 Reading comprehension

For this component students are required to read a text of approximately 1,500 words in length on a topic of general academic interest. The text is followed by a set of comprehension questions with a variety of question types including short answer, matching ideas, summarising, information transfer, gap-fill, true false and multiple-choice. Answers are marked using a specified marking scheme. Performance is reported as a series of band levels (from 4-9) indicating the level of support likely to be required in this skill area. A band score of 9 indicates that the student is highly competent at reading and comprehending academic texts. A band score of less than 7 indicates that s/he may be experiencing some difficulties in academic reading. A band score of less than 6 indicates that the student may be at risk of failure to due limited English reading abilities. The cut-offs between one band level and another were established at the University of Melbourne on the basis of consultation with ESL experts and academics from different disciplines.

There are three parallel versions of this component currently available. The new version was developed specifically for the University of Auckland context. Development of a subsequent version is also underway.

3.2.2 *Listening comprehension

This listening assessment is made up of a mini-lecture on a topic of general academic interest that does not assume high levels of background knowledge. The lecture is followed by questions measuring

students' ability to extract specific information as well as to follow the gist of the subject matter. The lecture itself is approximately 8 minutes long and is divided into 3 sections, with 10 - 12 items associated with each section. The passage is played once only but students are provided with contextualized information in written form to help them orient themselves to the topic. A range of response formats is used including true-false, multiple-choice, and short answer. Again, as for Reading comprehension, performance is reported as series of band levels indicating the level of support likely to be required in this skill area. A bandscore of less than 6 indicates that the students may be at risk due to limited ability to comprehend spoken discourse in an academic environment.

There are three parallel forms of this task currently available and a fourth is under development.

3.2.3 *Argumentative essay

This is a longer writing task than the one described above requiring students to argue a position on a topic of general academic interest drawing on information contained in a series of brief written stimuli. The essay is scored analytically with a 9-point scale for fluency, content and form respectively, to allow a diagnostic profile of participants' strengths and weaknesses in writing to be produced. Performance is reported at a series of band levels that indicate the nature of performance at each level and the level of support likely to be required. Descriptions of typical writing performance at each band level are provided in Appendix 1. There are four versions of this essay task currently available.

4 THE AUCKLAND PILOT

There were two separate phases of the Auckland pilot. The first involved the trialling of the new assessment components i.e. Text-editing, Report writing and Reading comprehension (Version 3) and was followed by item analysis and revisions (see 4.1.1. below). The second involved the administration of all existing and new (revised) components to two different groups of students enrolled at the University of Auckland who were reported to be experiencing difficulties with English. The purpose of this second trial was three-fold: a) to ascertain test properties with a larger population of students, b) to provide information about their English language needs with a view to considering the kinds of supports which could be put in place for these and future students identified as being at risk with their English (see 4.2.6. below) and c) to set thresholds on the screening component at or above which students could be exempted from the diagnostic assessment (see 4.2.7.).

4.1 Trial of the new assessment components

This trial took place in January and May 2001 and involved:

- a) students (N = 40) enrolled in various summer school courses in the Department of Applied Language Studies and Linguistics, with the majority drawn from Stage One credit courses in ESOL (English for speakers of other languages),
- b) students (N = 61) enrolled in a range of first semester courses within the Faculty of Arts. (The majority of these were enrolled in first year Philosophy subjects.)

Prior to the commencement of the assessment procedure, students were asked to complete a Background Information Questionnaire (Appendix 2) soliciting a variety of information about their English and other language experience. They were also given information about the DELNA research project and asked to sign a Consent Form in keeping with requirements of the Human Subject Ethics Committee.

The sample comprised 101 students in total. Of these 24% declared that English was their first language and 71% reported themselves as being from non-English speaking backgrounds. No language background information was available for the remaining 5%. About a third of the NESB students had spent 2 years or less in an English-speaking country. The remainder had been in New Zealand or another English-speaking environment for more than 2 years. Participants were all volunteers, and were offered the sum of \$30.00 in compensation for time spent, as well as feedback on their performance if they so desired. The trial took around two hours.

4.1.1 Item analysis

Following marking and data entry, Classical and/or Item Response Theory analyses were undertaken to investigate the properties of the objectively scored assessment components. Key results of these analyses are summarized below.

Text Editing

Data from all 101 participants was analysed using the ACER *Quest* software (Adams & Khoo, 1993) so that any misfitting items could be identified. Problems with 3 of the 73 items were identified and solved by using a different word or by moving the inserted word to another part of the line to avoid ambiguity. Item fit statistics were also used to assess the presence of a secondary speededness or time variable. If present it was expected that the items toward the end of the assessment task would have relatively larger values than those at the beginning. No such pattern was detected. The overall case reliability coefficient yielded by the *Quest* programme was 0.87, which was deemed acceptable.

Reading comprehension

Items from the new version of the Reading comprehension assessment were first analysed separately. A number of poorly discriminating items were identified on the basis of this analysis and either amended or removed from the pilot version. Second, test equating methods were undertaken whereby items with known properties from an earlier version (developed and validated in Melbourne) were anchored to those in the new reading task to establish their relative difficulty. Adjustments were then made to the new version, including the removal of a number of the easier items, to bring it into line with the old one in terms of both its overall difficulty and the spread of items across the ability range. It should be noted that the resultant assessment task was considerably shorter than the original one allowing for a reduction in the amount of time required to complete it. The analyses were rerun after the above modifications had been made yielding a case reliability estimate of 0.84, which was deemed satisfactory given the reduced length of the procedure.

Report writing

Writing scripts were sorted into rank order by two experienced ESL assessors and divided into 4 separate sets (following procedures advocated by Upshur & Turner, 1996). Benchmark scripts were then chosen and these formed the basis for drawing up a holistic description of typical performance at each level. These descriptors formed the basis of a rating scale that was used for marking all subsequent scripts. The reliability of these ratings was ascertained by correlating the scores assigned by the two raters concerned. The resultant coefficient was 0.85. The scripts that the raters disagreed about were then re-examined and an attempt was made to resolve any discrepancies by making adjustments to the rating scale descriptors as appropriate.

4.1.2 Participant feedback

Participants were asked to fill out a questionnaire following their completion of three components. Questions were designed to give information about the appropriateness of instructions, level of difficulty, time allocation, etc. (see Appendix 3). Students were also asked whether they considered each assessment task to be a good measure of their language ability.

Feedback was generally positive indicating that these new components were acceptable to the majority of the trial participants. Over 80% of students who completed the questionnaire considered the reading

and writing tasks good measures of their language ability. Students made a range of comments but in general these did not point to any consistent problem that they had experienced in completing any one assessment task. A number of native speakers of English found that the time allocation for the Writing tasks was too generous. In subsequent assessment trials writing tasks were sequenced before a break to give those who had completed them the chance to leave the room if they so wished. The Text-editing task was slightly less popular than the others with only 74% of students regarding it as a good measure of their ability. The main criticism from participants was that they were unable to finish the task. Since this was deliberately designed as a speed test, it was decided to signal this more clearly to the students by changing its title to *Speed reading* and warning them that many of them would be unable to finish the task within the allocated time.

4.2 Trial of the entire assessment battery

All six components of the assessment battery were trialled on two populations of students, those in the Bachelor of Business and International Management and those in the Wellesley foundation studies programme. It should be noted that the impetus for these pilot administrations came from the administrators of the programmes who perceived many of their students to be performing poorly in their courses due to difficulties with English. These samples may not therefore be representative of the wider University population. They nevertheless served as useful case studies which allowed us to work through the logistics of DELNA administration and reporting. We will report first on the overall quality of the assessment procedure as revealed in this more extensive trial. Information will then be provided about the English abilities of the students concerned.

4.2.1 Participants

BBIM group

Participants for this trial were drawn from students enrolled in the Business Information and Management Course offered at North Shore and Tamaki campuses. An initial proposal made to students failed to motivate anyone to participate in the trial. It was realised that the incentives offered (\$20 free Internet credit) were perceived to be inadequate. It was then proposed that students be offered \$40 remuneration (\$20 funded by the School of Business and Economics and \$20 from the Vice-Chancellor's Fund designated for DELNA development). On this occasion incentives were sufficient to ensure reasonable student participation (a free lunch was also provided to Tamaki students on completion of the assessment).

Assessment took place on May 28th and May 31st at Tamaki campus. A total of 85 students attended one of these two assessment sessions, 78% of the total number of students enrolled in the course. Assessment at the North Shore Campus took place on May 29th and 54 students attended this session (68% of the total number of students enrolled in the course). The total number of students participating in the trial was therefore 139.

Wellesley group

Students in the Wellesley programme were assessed on the trial over a 2-week period at the beginning of the second semester. The assessment tasks were administered in separate components on July 17th, 19th, 24th and 26th. Participation was voluntary but as assessment was scheduled during normal class time, most students elected to be part of the procedure. The course co-ordinator offered, in addition, an extension for one assignment to all students who volunteered to participate and attended all sessions. The scripts of a number of students who participated in the assessment were not however marked and included in the evaluation either because of late arrival to class (this posed a considerable problem throughout the trial) or because of sporadic attendance over the assessment period. A total of 67 students had results included in the trial evaluation, however only 30 of these were present to complete all assessment tasks.

4.2.2 Administration procedures

Background information and consent forms were obtained for each students as in the previous trial. Administration of all tasks took a total of 3 hours. Students in the BBIM trial were allowed a 5-minute break at the end of the first hour. Students were wherever possible seated so as to minimise the chances of their being able to see the answer sheets of fellow students; this however proved difficult at North Shore due to a higher turnout of students than originally anticipated. The unpredictable attendance and sporadic arrival of Wellesley programme students also made optimum seating arrangements at times difficult to achieve.

4.2.3 Rating procedures

Following the initial assessment trial three raters were chosen to attend two training sessions (each of one and a half-hour's duration) in order to learn how to score the Vocabulary, Speed-reading, Listening comprehension and Reading comprehension components. They received feedback on their marking of a set of benchmark scripts and were then given the rest of the scripts to mark. For benchmarking purposes the two sets of Writing scripts (report-writing and argumentative essay) were sent to the University of Melbourne (a *Universitas 21* member institution) where they were blind marked by a pool of 4 trained language assessors against a set of pre-specified rating criteria. The shorter Report writing task was scored holistically whereas the longer Argumentative essay task was scored analytically. All writing scripts were double marked to boost reliability and most participants were assessed by 4 different raters (2 for each rating task). Where there was disagreement between assessors, the more generous score was allocated to the candidate.

4.2.4 Quality of the assessment tasks

To ascertain that the measures developed were reliable when administered to this larger and more disparate population of students, case reliability analyses were performed on the combined data set of Wellesley and BBIM students for all 6 components. Results are reported in Table 1 below.

DELNA components	Case reliability estimates
Vocabulary	0.87
Speed reading/Text editing	0.88
Listening comprehension	0.82
Reading comprehension	0.83

Table 1: Case reliability estimates for objectively scored components

Reliability levels are somewhat lower for listening and reading comprehension than when the procedure was initially trialled and validated in Melbourne. It was found that one of the factors contributing to this somewhat lower level of reliability was the presence of a substantial number of ESB students in the sample, whose responses to some items may have differed from those of the NESB participants because of the very different nature of their language background and learning experience. A bias analysis (using the ACER *Quest* software) was undertaken and showed that 4 of the Listening comprehension and 7 of the Reading comprehension items functioned differentially for native and non native students with the majority of “biased” items⁴ favouring the NESB students (for further details see Elder, McNamara & Congdon, forthcoming). In other words, a diagnostic profile based on typical performance of NESB students may not provide an accurate picture of the other group’s performance. The sources of this differential item functioning are currently being examined and further modifications may be made to the answer key and or assessment items to accommodate these differences. It is hoped that these modifications, coupled with more rigorous rater training in the application of the marking guide, will boost the reliability of these components. It is also worth mentioning that the biasing effect of the native speaker presence is likely to be mitigated once the initial screening on the Speed-reading and Vocabulary components (see 4.2.7. below) has been applied. It is anticipated that this will result in many of the ESB students being exempted from the Listening, Reading and Writing components resulting in a more homogeneous population.

As for the Report and Argumentative writing tasks, interrater reliability checks between different pairs of assessors were conducted and were found to range from 0.79 and 0.96. The 0.79 correlation, which was between two sets of ratings on the report writing task, falls below what can be regarded as acceptable. Feedback from the assessors in Melbourne has led to minor revisions to the wording of instructions on the Report writing task and to revisions to the assessment scale to allow for 5 rather than 4 levels of performance. The revised scale is currently being applied to existing scripts and it is hoped that will allow for finer distinctions and therefore more accurate estimates of ability at the preliminary screening stage. A face-to-face rater training session on the use of this scale will be conducted in due course and it is anticipated that this will have the effect of boosting reliability on this

⁴ We are using the term ‘bias’ in the measurement sense, namely: “If an item’s estimated difficulty is significantly greater when calibrated on one subgroup than when calibrated on the other, the item is considered ‘biased’ with respect to those two” (Masters 1988:17). In this case the subgroups of interest are the ESB and NESB learners.

task. (Note that a self-training package containing a small sample of benchmark scripts was all that was available for the Melbourne raters.)

A bias analysis was also conducted on the writing tasks using the ACER *ConQuest* software (Wu, Adams & Wilson, 1998). While the shorter, holistically-scored Report writing task functioned similarly for the ESB and NESB students, the longer, Argumentative writing revealed significant bias with respect to two of the three analytic rating categories used to assess performance. More precisely, ESB students at a given score level were more likely to perform significantly better in terms of the formal accuracy (and, to a lesser extent, fluency) of their writing than NESB students with the same overall score (for further details see Elder *et al* forthcoming). This may be an indication that one or more raters are applying different standards when assessing scripts produced by native and non-native students. While this is not surprising, in practical terms this is likely to limit the validity of the diagnostic profiles provided for each group. Consideration is therefore being given to developing a separate rating scale for ESB students, with descriptors at each score level formulated differently from those provided in Appendix 1. Raters will need to be trained in applying the scale that best fits the performance at hand. The different pattern of ESB and NESB student performance also has resource implications, indicating the need for different kinds of support for members of each group.

Correlational analyses were also performed to determine the relationship between the various assessment components. Results are reported in Table 2 below.

	Vocab.	Speed read.	Listen	Read	Report writing	Argum. writing	Speed read. & vocab	Overall
Vocab.		.592**	.787**	.689**	.503**	.559**	.709**	.742**
Speed read.	.592**		.706**	.663**	.544**	.618**	.988**	.772**
Listen.	.787**	.706**		.749**	.509**	.571**	.759**	.868**
Read.	.689**	.663**	.749**		.465**	.563**	.712**	.859**
Speed read & vocab	.709**	.988**	.759**	.712**	.583**	.649**		.818**
Report writing	.503**	.544**	.509**	.465**		.473**	.583**	.581**
Argum. Writing	.559**	.618**	.571**	.563**	.473**		.649**	.820**
Overall	.742**	.772**	.868**	.859**	.581**	.820**	.818**	

** significant at the 0.01 level (2-tailed)

Table 2: Correlations between the various assessment components

The correlations between the components of the procedure give information that is pertinent to the practical issue of whether all components are making independent contributions to the measurement of

candidate ability. The relationship between the Reading comprehension, Listening comprehension and the two Writing components are moderate, suggesting that although each is related to the other, they are all measuring separate abilities and therefore make separate contributions to students' ability profile. It is interesting (if somewhat unexpected) to observe that the correlation between the Report writing and Argumentative essay tasks is only 0.47, substantially lower in fact than the correlation between Listening and Reading comprehension (0.75). This may have to do with the different nature of each task - one being more factually based and the other requiring argumentation - but also with the nature of the scales, with the former being holistic and somewhat shorter than the latter (4 points as opposed to 6). The proposed adjustments to the scale (see above) may have the effect of strengthening the correlation between the two writing tasks and this is currently being investigated. At this stage however it would appear that the two tasks are eliciting different skills from students and that removal of one of these from the procedure, in the interests of efficiency, is not warranted.

Vocabulary, as anticipated, is significantly related to performance on all other assessment components and the relationship is strongest for Listening ($r = 0.79$) Speed-reading is also a robust predictor of performance on the other assessment components and correlates slightly more strongly with writing performance (0.54 for the report writing task and 0.62 for the argumentative essay) and with the overall score (0.77) than does Vocabulary. When Speed-reading and Vocabulary scores are combined, the correlation with overall performance is 0.82. This finding is reassuring and suggests that it should be feasible to use these two components to exempt substantial numbers of students from being assessed on the full assessment battery. The issue of where to set the thresholds for exemptions is taken up later in this report.

4.2.5 Performance of the trial candidature

Descriptive statistics for BBIM and Wellesley student performance across all components of DELNA are shown in Tables 3 & 4 below.

	Vocabulary (Maximum possible 27)			Speed reading (Maximum possible 73)			Report writing (Maximum possible 4)		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
WP	54	23.44	3.15	52	32.38	16.70	65	2.98	.72
BBIM	139	23.07	4.26	139	35.44	19.14	139	2.88	.86

Table 3: Performance of the trial population on DELNA Screening component

	Listening (Maximum possible 39)			Reading (Maximum possible 41)			Argument writing (Maximum possible 9)		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
WP	51	27.37	7.21	42	23.88	7.72	47	6.11	0.96
BBIM	139	29.30	7.41	139	27.91	8.20	138	6.47	1.31

Table 4: Performance of the trial population on DELNA diagnostic component

The two groups perform at similar levels on the Vocabulary and Report writing components, but the BBIM students perform somewhat better than the others on Speed-reading, Listening and Reading. It should be noted however that there is a greater proportion NESB students in the BBIM group (56%) than amongst the Wellesley students (48%) and this may have depressed the overall scores for the former group.

A breakdown of performance of each group according to language background is presented in Tables 5 to 7 below. For the BBIM students, the NESB students have been divided into two sub-groups: students who have spent more than 2 years in New Zealand or another English speaking country and recent arrivals who have been in the country for 2 years or less. There were insufficient numbers of recent arrivals in the Wellesley group to warrant this subdivision. Worthy of mention is the fact that of the 32 Wellesley students who identified their first language as being other than English, 10 indicated that their first language was a language of Polynesian origin (including Maori).

	Vocabulary (27)		Speed Reading (73)		Report Writing (4)		Listening (39)		Reading (41)		Argument writing (9)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ESB <i>N</i> = 61	25.8	1.2	49.1	15.6	3.5	.67	34.6	3.3	32.5	5.2	7.4	.89
NESB (more than 2 yrs) <i>N</i> = 34	22.2	3.6	27.7	15.1	2.5	.70	28.2	5.3	26.5	7.4	5.9	1.2
NESB (less than 2 yrs) <i>N</i> = 44	18.9	4.9	20.9	12.4	2.2	.70	21.2	7.3	21.6	8.8	5.9	1.2

Table 5: Performance of BBIM students on DELNA by language background

	Vocabulary (27)		Speed Reading (73)		Report Writing (4)		Listening (39)		Reading (41)		Argument writing (9)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ESB N= 35*	24.0	1.8	37.8	16.0	3.3	.65	29.7	5.2	27.6	6.1	6.5	.71
NESB N= 32*	20.8	3.5	26.1	15.6	2.7	.67	24.8	8.3	18.4	6.5	5.7	1.1

*These numbers are only indicative, since the numbers taking each component differ and the proportion of NESB to ESB English speaking students fluctuates accordingly.

Table 6: Performance of Wellesley students on DELNA by language background

Results indicate, as might have been expected, that the ESB students perform better on all DELNA components than the NESB students. In the case of the BBIM sample, the means are, again predictably, lower for recent arrivals than for students who have spent more than 2 years in an English-speaking country, except in the case of the Argumentative writing task where the two NESB groups perform on a par with one another.

4.2.6 Profile of English language needs

The above information, while allowing comparisons across groups and sub-groups within the candidature, has limited utility with regard to identifying the language support needs of the target population. The figures below show the distribution of the BBIM and Wellesley groups across 6 different bands of ability, which are linked to the level of English language support required. These ability bands have been reported for Listening and Reading comprehension and for the Argumentative writing task only, since the other assessment components have been developed for preliminary screening rather than diagnosis.

BBIM students

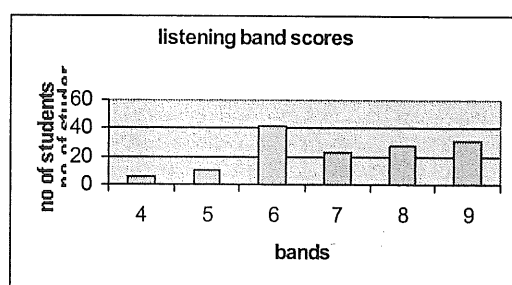


Figure 1: Distribution of BBIM students across Listening bands

Band	4: at severe risk, urgent need of language support	(6 students / 4%)
	5: at risk, needs extensive language support	(10 students / 7%)
	6: concurrent English language support advisable	(41 students / 30%)
	7: may benefit from further English	(23 students / 17%)
	8 & 9: unlikely to require English support, competent listener	(59 students / 42%)

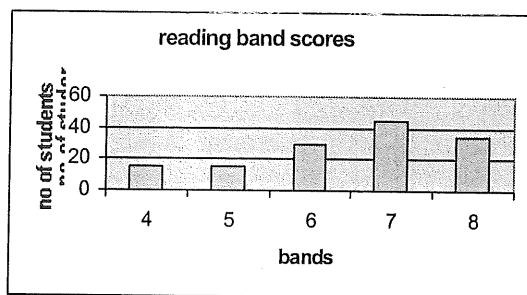


Figure 2: Distribution of BBIM students across Reading bands

Band	4: at severe risk, urgent need of support	(3 students / 7%)
	5: at severe risk, needs extensive support	(11 students / 26%)
	6: concurrent English language support advisable	(16 students / 39%)
	7: may benefit from further English	(9 students / 21%)
	8 & 9: unlikely to require English support, competent reader	(3 students / 7%)

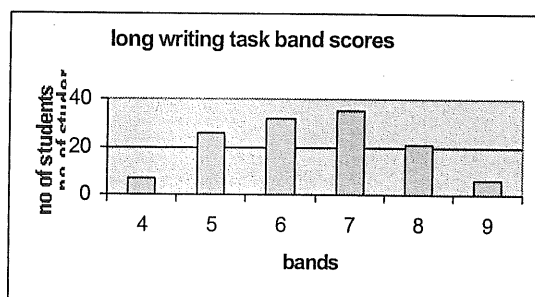


Figure 3: Distribution of BBIM students across Writing bands

Band	4: at severe risk, urgent need of support	(7 students / 5%)
	5: at risk, needs extensive support	(26 students / 19%)
	6: concurrent English language support advisable	(32 students / 23%)
	7: may benefit from further English	(35 students / 26%)
	8 & 9: unlikely to require English support, competent reader	(37 students / 27%)

See Appendix 1 for profiles of writing ability at each band level.

Of the 138 BBIM students assessed an overall total of 86 (62%) needed support with their English (i.e. their scores on one or other of the assessment components fell within the Band 4-6 range). 26 of these needed help with one skill area, 25 with 2 skill areas and 35 with all 3 skill areas.

In terms of urgency of need it should be noted that 48 of the 86 students could be regarded as at risk or severe risk in their academic studies due to limited English skills (Band 4 or Band 5). The vast majority (N=46) of this latter group identified themselves as being from non-English speaking backgrounds.

Wellesley students:

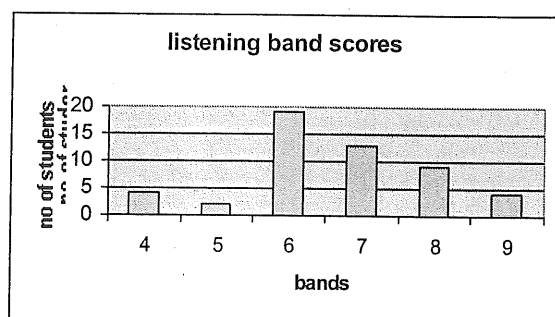


Figure 4: Distribution of Wellesley students across Listening bands

Band	4: at severe risk, urgent need of support	(4 students / 8%)
	5: at severe risk, needs extensive support	(2 students / 4%)
	6: needs concurrent English language support	(19 students / 37%)
	7: would benefit from further English	(13 students / 25%)
	8 & 9: unlikely to require English support, competent listener	(13 students / 26%)

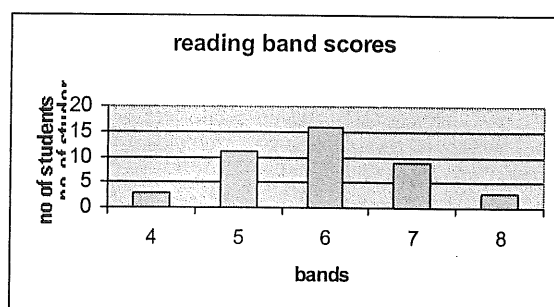


Figure 5: Distribution of Wellesley students across Reading bands

Band	4: at severe risk, urgent need of support	(3 students / 7%)
	5: at severe risk, needs extensive support	(11 students / 26%)
	6: concurrent English language support advisable	(16 students / 39%)
	7: would benefit from further English	(9 students / 21%)
	8: unlikely to require English support, competent reader	(3 students / 7%)

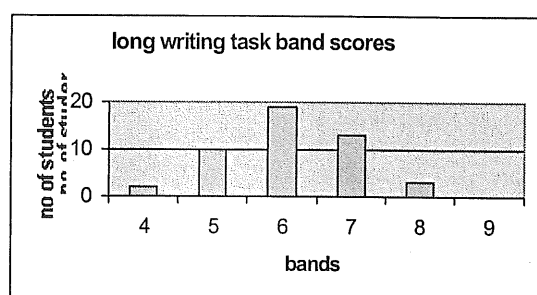


Fig 6: Distribution of Wellesley students across Writing bands

Band	4: at severe risk, urgent need of support	(2 students / 4%)
	5: at risk, needs extensive support	(10 students / 21%)
	6: concurrent English language support advisable	(19 students / 41%)
	7: may benefit from further writing instruction	(13 students / 28%)
	8 & 9: unlikely to require English support, competent writer	(3 students / 6%)

In interpreting the results for students in the Wellesley programme it is to be noted that more than half of the students assessed (37) did not complete all assessment tasks, even though the procedure was administered during class time.

Of the 67 students assessed 74% would clearly benefit from English language support on one or more of the three language skills. Just under half of these (29 students) needed urgent language support (Bands 4-5 on any one of the assessment components). 18 (i.e. 69%) of this high-risk group identified themselves as being from non English-speaking backgrounds.

28 students were at the Band 6 level, indicating that they would well advised to undertake concurrent support with their English while pursuing their academic studies. 14 of these were native speakers of English and 3 were from Maori or Pacific island backgrounds.

4.2.7 Initial proficiency screening

The moderately strong relationship between scores on the Vocabulary and Speed reading task reported in Section 4.2.4 above suggests that it is feasible to use these two components as a screening device by setting a threshold at or above which some students can be exempted from being assessed on all subsequent assessment tasks. A score of 75 across the two tasks has been tentatively set as a cut-off. Had this threshold been applied to the BBIM data, 25% of students would have been exempt from assessment on the more labour intensive Reading, Listening and Writing components, most of them native speakers of English. Only four students in this exempt group scored lower than Band 7 or 8 on any of the Listening, Reading and Writing components. Amongst the Wellesley population, however, the application of this threshold would allow us to exempt only 4 (15%) of the candidature from being assessed on the remaining components. If the threshold were set lower (at 70) we would still capture the vast majority of high-risk students. However such a threshold would result in the loss of diagnostic information about some students at the Band 6 level who might benefit from additional English support in one or more skill areas. A compromise solution would be to require satisfactory performance on the shorter Report writing task before exempting those in the 70-75 score range from further diagnostic assessment.

While the above thresholds will need to be refined and confidence levels established with a larger data set, the idea of an initial screening procedure appears to be feasible and should result in considerable savings in the costs of marking and the generating of report profiles. It is nevertheless recommended that all participants be required to sit for all components regardless of whether their scripts are marked. The logistics of recalling some students for a second round of assessment would be unwieldy and likely to create undue anxiety. For those who achieve a score at or above the requisite threshold, a brief statement to the effect that performance is satisfactory could be issued to participants at the same time as the fuller diagnostic reports are generated for the remaining students.

4.2.8 Relationship between DELNA assessment results and University grades

Evidence for the utility/relevance of the English diagnostic assessment procedure was sought by correlating performance on DELNA with participants' academic performance as measured by grades obtained for various academic subjects. A moderate relationship could be taken as evidence that language is a significant factor in student performance at the university and therefore that some kind of intervention strategy for students with limited English may be appropriate.

BBIM students: DELNA assessment results were compared with two measures of University performance. The first of these was the students' GPA for the first semester of 2001 and the second was their grade on Management 191, a paper that was estimated to have a more substantial writing component than the others. These two measures were correlated with the sum of the students' total Listening, Reading and Argumentative writing band scores. Results (see Table 7 below) show a stronger relationship between the various components of English language ability and academic performance than is usually reported in comparable predictive validation studies (where a correlation of around 0.3 is the norm). Language proficiency here accounts for almost a third of the GPA variance and over 50% of the variance in Management grades. Interestingly, Listening is the best predictor of performance on the Management paper and Reading has a stronger relationship than the other skills with overall GPA.

	Listening	Reading	Writing	Total (L, R & W)
Management 191 <i>N=139</i>	.686	.664	.547	.721**
GPA <i>N=139</i>	.525	.600	.322	.556**

*significant at the 0.01 level (2-tailed).

Table 7: Correlation between DELNA results and Semester 1 academic grades on the BBIM

Wellesley programme students: DELNA assessment grades were compared with the English language programme end of Semester marks. Again, this measure was correlated with the sum of the students' total Listening, Reading and Writing band scores. Results are consistent with what was reported for the BBIM group and confirm that English language proficiency is an important factor in academic performance (see Table 8 below). This is less surprising however, given that the academic results are for English which one would expect to make heavy demands on language proficiency. These results should nevertheless be interpreted with caution given the small size of the student sample and the fluctuation in numbers across the various DELNA components.

	Listening	Reading	Writing	Total (L, R & W)
Wellesley English	.449**	.385*	.383**	.532**

* significant at the 0.05 level (2-tailed)

** significant at the 0.01 level (2-tailed)

Table 8: Correlation between DELNA results and mid-semester English academic grades on the Wellesley programme

One factor that may be contributing to the higher than usual correlations between language proficiency and academic performance is the fact that the DELNA was administered at the same time or very close in time to when academic grades were assigned. In most predictive validation studies there is a greater gap in time between the gathering of language proficiency information and the issuing of grades, by which stage students' language proficiency could be expected to have improved (Ferguson & White, 1998). It may also be that students are being accepted into these particular academic courses with lower levels of English proficiency than has been the norm at other academic institutions where predictive validation studies have been conducted. There is generally agreed to be a threshold of proficiency (e.g. see Criper & Davies, 1988; Elder, 1993; Graham, 1987) below which English is a powerful factor in academic performance. Above that threshold it seems that a range of factors, including language proficiency, may come into play. In the case of the BBIM group, the high-risk level appears to be at around the Band 5 level or below. 31 students had a band score of 5 and more than half of these had a GPA of less than 3. 15 students had a Band score level of 4 and 11 of these had a GPA of less than 2.

The same threshold seems to operate for the Wellesley group. Of the 17 students with a band score of 5, all scored a B- (65%) or less for English and two of them failed the first semester of their English course. All 5 students with band score of 4 received a Fail grade for English.

4.2.9 Reporting and uptake

Summary information about the performance of the trial candidature was offered to the staff of the participating programmes in the form of a brief report. Individual results were not included. These were reported individually to participants in the form of a profile issued via email or post. On the profile band scores were converted to percentages (i.e. a band score of 4 was reported as 40) to avoid any confusion between DELNA and IELTS, which is also reported on a 6-point scale. Students were given their results for Listening, Reading and the Writing, together with information about the range of scores for each of these assessments. Students were advised that if they scored 40 or 50 on any one of these assessments that their English skills were severely limited and that this could affect their future performance at the University. They were also urged to contact one of their lecturers for advice about how to access English language support. In addition they were also offered one-on-one support with their coursework essays from an ESL teacher who was conducting research into students' writing needs. Anonymity was guaranteed.

Interestingly, no student from the BBIM approached the lecturer although a number of them subsequently contacted the volunteering ESL teacher seeking help with particular essays. Information is not yet available for students on the Wellesley programme but it is likely that at this late stage of the academic year uptake will not be immediate and that NESB students responding to the advice will do so by enrolling in one of the summer school ESOL credit courses offered by the Department of Applied Language Studies and Linguistics.

Arrangements have already been made for administering DELNA to students in next year's Wellesley programme. This information will be used to create separate streams in that programme with the more proficient group taking the regular English programme and those with limited proficiency taking a parallel ESOL credit course which will be tailored to their specific language needs. A policy for the ESB students who also need additional support with English has yet to be established.

In the case of BBIM students, there are plans to administer the procedure next year but as yet no clear policy as to how the results will be used. The question of what help can be provided for students who have no room in their Bachelor degree programme for credit courses in English (whether for ESB or NESB students) remains unresolved and will be discussed further in Section 5 below.

The generally poor uptake of advice amongst the BBIM students participating in the pilot raises the issue of the sensitivity of the information provided by DELNA (which the students may well have seen as jeopardising their progress at the university in spite of clear assurances to the contrary). This issue will also be revisited in Section 5.

5 DELNA DELIVERY

The university's policy is to deliver DELNA university-wide to all students other than those meeting specified exemption criteria. This raises a number of logistical issues that will be discussed below.

5.1 Proposed exemptions

In May 2001 Professor Ellis proposed a number of criteria to be used as a basis for exempting students from the requirement to sit for DELNA. Those who did not meet these criteria were deemed to be potentially at risk in so far as they lacked sufficient evidence of English language proficiency. The proposed exemptions would apply to undergraduates who meet any ONE of the following requirements described below:

- *University Entrance Bursaries (UEB) and Scholarship or Education Scholarship Trust (EST):* any student with a score of 65% or higher in the following subjects: English; History; Classical Studies; History of Art.
- *Sixth Form Certificate:* any student with a level 1 or level 2 pass in the following subjects: English; History; Classical Studies; History of Art.
- *International Baccalaureate:* any student who has been awarded an IB Diploma, which necessarily includes satisfactory completion of the Extended Essay.
- *Examinations for NESB students:*
 - a. IELTS Examination: any student with 6.0 or more in the IELTS examination, including a minimum of 6.0 in the Writing Section;
 - b. TOEFL: any student with a total score of 560+ with no section scores below 56 (or 220+ on the new computer-based TOEFL, with no section scores below 22);
 - c. Cambridge Certificate in Advanced English (CAE): any student with a B grade;
 - d. Certificate of Proficiency in English (CPE): any student with a C grade.
- *Wellesley Programme:* any student who achieves a B or higher on the English examination.

The BBIM student group is a good example of a population that has entered the university with limited exposure to school subjects that can be regarded as requiring highly developed skills in the English language. Of the 139 students who sat for DELNA, only 37 had sat Bursary English, 9 Bursary Classics and 5 Bursary History. Twenty-four students had passed English, 7 had passed Classics and 5 had passed History. Moreover, grades among these students were generally low, with only 13 students gaining 60% or more in English, 2 in Classics and 1 in History.

Background information supplied by the Enrolment and Admissions Office indicates that if the proposed exemption categories had been applied to this particular group of undergraduates, 32 of the participants (23%) would have been exempted from participating in the DELNA assessment. Of these students all but two scored above bands level 4 and 5 on the Listening, Reading and Writing components, indicating that they were not in urgent need of English language support. Twelve students

scored at the Band 6 level on one or more of these tasks indicating that they may have benefited from further training in academic English but that their needs were not urgent. This can be taken as tentative support for the proposed exemptions policy, which appears to capture most, if not all, of the high-risk students.

As far as the Wellesley programme is concerned, it should be noted that all 4 of the students who achieved B+ or higher on the English exam performed at satisfactory levels (70 or above) on all components of DELNA.

5.2 Feasibility

5.2.1 Identification of the target candidature

There are however real problems in implementing the proposed strategy. One has to do with the difficulty of accessing the relevant information about the students concerned and hence identifying the target candidature. An attempt to estimate the numbers of students who would have sat DELNA, had this exemptions policy been applied to the whole population of incoming undergraduates in 2001, has proved very difficult.

It was possible to establish that there are approximately 3000 students who sat Bursary in 2000 and did not achieve the requisite 65% on one or other of UEBS English, History, Classical Studies or History of Art. These students could be readily identified at the stage of matching the Bursary tape to admissions information. It was also established that there would be no difficulty in accessing information about Wellesley students' English scores.

However, the advice from the Admissions office is that:

- Sixth Form Certificate results are not recorded in *nDeva*, except for 3 or 4 specific programmes. The certificate is scanned as an image only;
- International Baccalaureate is used as an entrance qualification but not recorded specifically in *nDeva*;
- IELTS and TOEFL are recorded on the admissions form as having been sat, but test scores are not recorded in *nDeva* in a consistent fashion, so the overall level requirement and the information about performance across sub skills is not accessible;
- The Cambridge Certificate in Advanced English and Certificate of Proficiency in English are not recorded in *nDeva*.

This means that it is not possible to write a query or set of queries in *nDeva* that identifies the target group with the exemptions removed. The possibility of capturing all relevant students does not currently seem feasible and suggests that the proposed list of exemptions may need to be modified, and in the longer term, better mechanisms for identifying the target student population will need to be

established. There is no possibility of instituting DELNA as a mandatory requirement until more watertight systems for student identification can be put in place.

5.2.2 Timing

The second logistical problem has to do with timing of the DELNA administrations. Applications for admission to the university can be made in early December but in practice continue into the early weeks of Semester 1 in the following year. The matching of Bursary information to Admissions information occurs in mid January. The task of a) identifying relevant students b) notifying them of DELNA dates and procedures, c) marking and delivery of DELNA results and d) provision of advice regarding a suitable course of action is likely to take at least two weeks (one week for identification and notification, one week for administration of DELNA and turn-around of results). While some students may be able to be identified early, information about which students are at risk may in many cases come too late for them to enrol in English enhancement courses. Many students identified in first semester may therefore not be able to act on advice received until second semester. Successive administrations will need to be conducted over the course of the year so that students who were unavailable to sit DELNA in late January as well as those who enrol in second semester can be assessed. Although the total number of administrations will need to be restricted, it is recommended that a number of alternative dates be fixed for late January and mid February after offers have gone out. Further sessions will need to be held during the mid-semester break and again in June prior to the start of second semester.

5.2.3 Reporting

A key issue in the successful delivery of DELNA and subsequent uptake is how the procedure is described to incoming undergraduates and how the results are reported to them. It is imperative that DELNA information does not end up being used to exclude students from enrolling in particular degree courses. Clear protocols will need to be developed for reporting purposes, with individualised profiles issued confidentially to the students, and summary information about students in the various needs categories supplied only to designated individuals within each Faculty who have been briefed about the purpose of DELNA.

5.2.4 English language support

In addition the larger issues of a) what constitutes suitable intervention for at risk students, b) how students receive advice on the various possibilities available to them and c) how to ensure uptake of this advice need to be addressed. It is currently proposed to list all current sources of support on a Web page which students can access on receiving their results, with particular courses highlighted as being useful for those experiencing difficulties with particular skill areas and distinctions made between courses designated for ESB and NESB students (e.g. ENG WRIT 101, ENGLISH 121 [for ESB], ESOL 101, ESOL 102, ESOL 201 [for NESB]). The current range of offerings is, however, unlikely to

be sufficient to meet the demand generated by DELNA, particularly as regards NESB students. Additional streams within existing courses as well as entirely new courses will need to be mounted to meet this need. A policy for linking information gleaned from DELNA to available course offerings has yet to be developed. Moreover, it should be noted that many Bachelor degrees offered by the University do not have any space for English writing electives, and that students may be reluctant or unable to take any of the non-credit courses currently available (e.g. ESOL 10E, ESOL 20E), due to timetable constraints and/or other logistical difficulties, as well as to the lack of any incentive associated with enrolment in such courses.

One possible solution to this situation would be to allow students to gain additional credits by taking one or other English option as a Certificate of Proficiency or as part of a proposed Certificate of Arts (ESOL). The English Language Self Access Centre (within the Department of Applied Language Studies and Linguistics) is a further source of support for at risk students from non English-speaking backgrounds but is unlikely to be sufficient for those with very low levels of proficiency (Band 4 & 5) who may need intensive input from a language teaching specialist which targets their particular areas of weakness. Short courses offered by the Student Learning Centre may also be helpful to students who need to be familiarised with macro-level discourse conventions of various kinds of academic writing.

6 RECOMMENDATIONS

The following recommendations are offered in the interests of expediting the delivery of DELNA in 2002 and beyond, and ensuring that the university is in a position to cater for the needs it identifies.

Recommendation 1

A DELNA administrator should be appointed as a matter of urgency, to establish procedures for administering DELNA in 2002 and to coordinate the process of DELNA delivery in 2002 and beyond. The administrator, working under the direction of a Project Manager, and with clerical assistance, would take responsibility for the following:

- establishment of mechanisms for identifying and contacting the target candidature;
- provision of information about DELNA to individuals and Faculties;
- identification and training of a large pool of raters;
- development of protocols for confidential reporting of performance to individual students and for the provision of aggregate performance data to the relevant Faculty;
- all aspects of DELNA administration including delivery of DELNA results;
- provision of advice to students regarding suitable English enhancement options;
- DELNA maintenance, including the on-going development and trialling of new parallel versions of each assessment component.

Recommendation 2

A university-wide short and long-term strategy plan for dealing with English language needs identified by DELNA should be developed in consultation with representatives from all Faculties, with particular attention paid to those at high risk of failure due to limited English skills.

Short-term strategies could include:

- directing students to available credit and non credit English enhancement courses;
- professional development for academic staff in ways of accommodating the needs students with a limited command of academic English within the context of mainstream lectures and tutorials⁵;
- team teaching on the part of, say, ESL and subject specialists lecturers with lectures offered by the former and tutorial assistance offered by the latter to ensure that the language demands of particular courses and assignments can be dealt with by the students concerned;
- one-on-one support for individual students as required.

⁵ As Zamel (1998) amongst others have argued, devising strategies to accommodate such students' needs should not be seen as a lowering of standards, but rather as a means of revitalizing higher education and improving the quality of teaching for all students.

Longer-term strategies could include a review of current degree requirements with a view to creating space for English enhancement courses within all university programmes. Until such electives are uniformly available to students university-wide, the impact of DELNA and any subsequent English language intervention is likely to be very limited.

Recommendation 3

Given current uncertainty about when and how DELNA will be administered, and what action should be taken to deal with the needs it identifies, it is recommended that DELNA be offered, in the first instance, to students declaring on the Admissions Form that they do not have English as a first language. These NESB students appear, on the basis of feedback from the assessment trials, to be those most in need of English language support. 2001 enrolment figures suggest that 2264 such students applied for entry to the University of Auckland in 2001. We would urge that all such students be required to take DELNA unless they can produce evidence of an IELTS score of 6.5 or a score of 65% or higher in the following subjects: English; History; Classical Studies or History of Art.

Recommendation 4

In addition to the above, special administrations could be set up for those Departments who are concerned about the English needs of their students and have set in place some mechanisms for a) identifying relevant students and b) dealing with such needs⁶.

Recommendation 5

Finally, the quality and consistency of information recorded on *nDeva* regarding the English language qualifications of applicants to the university needs to be improved before the beginning of enrolment in 2003. Until information about 6th Form Certificate, International Baccalaureate, IELTS, TOEFL, Cambridge Certificate and Certificate of Proficiency is readily retrievable, it will be impossible to implement DELNA on a mandatory basis university-wide and, accordingly, to address effectively the English language problems of the University of Auckland student population which emerge clearly from this report.

⁶ Such an arrangement has already been made for the 2002 intake for the Wellesley programme.

7 REFERENCES

- Adams, R.J., & S.T. Khoo (1993). Quest: The Interactive Test Analysis System [computer program]. Camberwell, Victoria Australian Council for Educational Research.
- Alderson, J.C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Adams, R. & S. Khoo (1993) *QUEST: The interactive test analysis system*. Victoria: Australian Council for Educational Research.
- Beglar, D. & A. Hunt (1999). Revising and validating the 2000 word level and the university word level vocabulary tests. *Language Testing* 16: 131-162.
- Clapham, C. (1996). *The development of IELTS: a study of the effect of background knowledge on reading comprehension*. Cambridge: Cambridge University Press.
- Criper, C. & A. Davies (1988). *ELTS validation report*. London: The British Council, Cambridge: University of Cambridge Local Examinations Syndicate.
- Davies, A. (1975). Two tests of speeded reading. In R.L. Jones & B. Spolsky (eds.) *Testing language proficiency*. Washington DC: Center for Applied Linguistics.
- Davies, A. (1989). *Testing reading speed through text retrieval*. In C.N. Candlin & T.F. McNamara (eds.) *Language Learning and Community*. Sydney, NSW: NCELTR
- Elder, C. (1993). Language proficiency as a predictor of performance in teacher education. *Melbourne Papers in Language Testing* 7: 15-63.
- Elder, C. T. McNamara & P. Congdon (forthcoming) Diagnosing the academic language proficiency of university entrants: can native and non native speakers be assessed in common? *Journal of Applied Measurement*.
- Ellis, R. (1998). Proposal for an English Language Proficiency Entrance Examination. Auckland, University of Auckland.
- Ferguson, G. & E. White (1998). A small-scale study of predictive validity. *Melbourne Papers in Language Testing* 7, 15-63
- Fulcher, G. (1997). An English language placement test: issues in reliability and validity. *Language Testing* 14, 2: 113-139
- Graham, J. G. (1987). English language proficiency and the prediction of academic success. *TESOL Quarterly* 21,3: 505-521.
- Gravatt, B., J.C. Richards & M. Lewis (1997). *Language needs in tertiary studies: ESL Students at the University of Auckland*. Institute of Language Teaching and Learning Occasional Papers, No. 10
- Hughes, A. (1988). 'Introducing a needs based test of English language proficiency into an English-medium university in Turkey. In Hughes A. (ed.) *ELT Documents 127: Testing English for University Study*. London: British Council.
- Language Testing Centre (1991). The University of Melbourne ESL test. Final Report. Melbourne: University of Melbourne.
- Loewen, S. & R. Ellis (2001). *The relationship of receptive vocabulary knowledge to academic success and learner beliefs*. Department of Applied Language Studies and Linguistics. Occasional Papers, 15.

- Moran, T. (1995). Report on the Sub-Committee on English Language and Entrance. University of Auckland.
- Nation, I.S.P. (1983). *Testing and teaching vocabulary. Guidelines*, 5, 12-25.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Rosenfield, M. S. Leung & P. K. Oltman (2001). *The Reading, Writing, Speaking and Listening Tasks important for academic success at the undergraduate and graduate levels*. TOEFL Monograph Series MS-21. Princeton, New Jersey: Educational Testing Service.
- Turner, C. & Upshur, J. (1996). Developing rating scales for the assessment of second language performance. In Wigglesworth, G. & Elder, C. (eds.) *The language testing cycle: from inception to washback*. ARAL Series S No. 13 Canberra: Australian National University.
- Wall, D., C. Clapham, & J.C. Alderson (1994). Evaluating a placement test. *Language Testing* 11: 321-344.
- Weir, C. J. (1983). Identifying the language problems of overseas students in tertiary education in the UK. Unpublished PhD thesis, Institute of Education, University of London.
- Wu, M., R.J. Adams & M.R. Wilson (1998). *ConQuest*. Camberwell, Victoria: Australian Council for Educational Research.
- Zamel, V. (1998). Strangers in Academia: The experiences of Faculty and ESL students across the curriculum. In Zamel, V. & R. Spack (eds.) *Negotiating Academic Literacies*. New Jersey: Lawrence Erlbaum.

APPENDIX 1: DIAGNOSTIC PROFILES OF WRITING PERFORMANCE

ARGUMENTATIVE WRITING

- Band 4:** This piece of writing is hard to interpret.
The point of view is not clear. Ideas and evidence are often confused and points made may not be relevant.
Few sentence patterns are used correctly. Vocabulary is limited and poor word choice often inhibits expression. Spelling errors are frequent.
- Band 5:** This answer causes problems of interpretation for the reader, although it displays an underlying coherence. Cohesive devices are inadequate, inappropriate or absent.
Ideas are put forward, but they are few. Some ideas may be irrelevant. There is probably a lack of evidence to support ideas.
A limited variety of sentence structures. Basic grammatical errors are noticed. Vocabulary is restricted. Spelling errors and poor word formation may cause strain for the reader.
- Band 6:** This is a mainly satisfactory piece of writing, although some strain may be caused by misuse or absence of cohesive devices.
The argument does not always progress logically and it is not always possible to distinguish ideas from evidence. Some points may appear irrelevant, and evidence may be lacking.
Sentence structures are generally adequate, although errors may occur frequently. Vocabulary limitations sometimes cause problems in expression of ideas. Some spelling errors are likely to occur.
- Band 7:** The essay generally reads fluently with only slight or occasional strain for the reader. Tone/style adopted may seem inappropriate for the task.
Arguments are put forward although at times the point of view may be unclear. Ideas are generally arranged logically. Ideas are generally relevant, but they may lack supporting evidence.
A satisfactory variety of sentence structures is used. Vocabulary is adequate, and generally used appropriately. Errors are few and unintrusive.
- Band 8:** A slight awkwardness of style is detectable. However, the message is easily followed. Tone/style may seem slightly inappropriate for the task.
There is a clear and logical progression of ideas, but examples or evidence may occasionally lack obvious relevance to the point being made.
There is a good variety of sentence structures. Vocabulary is wide and used appropriately. There are no significant errors in word formation or spelling.
- Band 9:** The essay is completely satisfactory.
The essay is clearly presented and well developed.
A wide variety of sentence structures is used, with no noticeable errors. Vocabulary is extensive and always used appropriately. There are no errors in morphology or spelling.

APPENDIX 2: BACKGROUND INFORMATION QUESTIONNAIRE

Please answer the following questions. All your answers will be kept entirely confidential..

1. What is your first language?
2. What other language(s) do you speak (not including English)?
3. What is your country of birth?
4. To which ethnic group or groups do you belong? (max 3)
.....
5. How long have you been living in New Zealand?
6. Have you ever lived in another English-speaking country? YES NO
(delete one)

If YES Which country? and How long?

7. Did you go to secondary school in NZ? YES NO (delete one)

If YES, for how long?

Did you go to secondary school in another English-speaking country? YES NO

If YES, Which country? and How long?

8. Have you sat any English language exams?

Bursary English	YES	NO	Score:	
TOEFL	YES	NO	Score:	
IELTS	YES	NO	Score:	
Other?	YES	NO	Please specify test:	and score:

The following information is confidential only and will not be given to any outside sources.

Family Name: Given Name:

Student ID:

We need the following in case we have to contact you. Once again this information is confidential.

Phone no: E-mail address:

APPENDIX 3: TEST-TAKER FEEDBACK QUESTIONNAIRE

Before you leave today, we would like you to tell us how you felt about the various parts of this assessment. Your responses will help us with various revisions to the procedure.

Please circle either SA (=Strongly agree) or A (=Agree) or D (=Disagree) or SD (=Strongly Disagree) in response to each of the statements below.

TEXT EDITING

The instructions were clear	SA	A	D	SD
There was not enough time	SA	A	D	SD
The text was too difficult	SA	A	D	SD
I thought this was a good measure of my language ability	SA	A	D	SD

Any other comments? _____

Please turn over the page

READING TASKS

The instructions were clear	SA	A	D	SD
There was not enough time	SA	A	D	SD
The texts were too difficult	SA	A	D	SD
The topics were appropriate for university students	SA	A	D	SD
There was a good range of items on these tasks	SA	A	D	SD
I thought the reading tasks were a good measure of my academic reading ability	SA	A	D	SD

Any other comments about one of the reading texts or items?

Please turn over the page

WRITING TASK

The instructions were clear	SA	A	D	SD
There was not enough time	SA	A	D	SD
The table was difficult to understand	SA	A	D	SD
The topics were appropriate for university students	SA	A	D	SD
I thought this was a good measure of my academic writing ability	SA	A	D	SD

Any other comments?

Thank you very much for helping us with the trialling of this procedure.