

Implementing NETPAW's diagnostic test of English proficiency in Australia: A case study¹

Shirley O'Neill
University of Southern Queensland
oneills@usq.edu.au

Abstract

This research is a small case study that explores the application of the National English Test in Proficiency for All on the Web (NETPAW) with international tertiary level students in Australia. It reports the results of three students' use of the tests of listening and reading and their opinion in the light of their previous English proficiency testing experiences. While small scale and exploratory it represents the first step towards the design of a larger study that will investigate both the usefulness of NETPAW for testing the knowledge and proficiency of a wider group of candidates who may wish to take advantage of being able to take a test of English proficiency online and also the views of educators who have a stake in being able to assess their students' English and collect diagnostic data to improve students' learning and their pedagogical approach. NETPAW is considered in terms of perceived interactivity, practicality, authenticity and validity (Bachman & Palmer, 2000) within the limitations of the exploratory nature of the study. The diagnostic nature of the test is explored through comparison with relevant aspects of a similar diagnostic test of English as a second language focusing on skill/item demands. All-in-all it was concluded that NETPAW offers a valuable alternative that is user-friendly through its availability online and its interactive assessment experiences. It has potentially multiple uses and international applicability for English language assessment and the teaching of English. This includes its diagnostic application. The fact that the NETPAW levels relate to the Common European Framework of Reference for Languages: Learning, Teaching and Assessment that describes learner proficiency at each level within a hierarchy for European languages is a further advantage since this assists in making links between NETPAW levels and other ESL/EFL scales and assessments.

Key words: language testing, EFL, ESL, English literacy, web-based language testing, diagnostic tests of English, language assessment scales.

Introduction

As the world wide web becomes the virtual global social world as well as the key information, communication and knowledge source the ability to use English to participate is more than ever paramount across the world. This has ensured the continued and unprecedented growth in the teaching of English and the inevitability of the need for learners, teachers and other stakeholders to be able to "measure" English proficiency. So in the milieu of this 21st century context the ability of information communication technology to positively impact on languages pedagogy and assessment emerges as an agent of potential dramatic change. Learners of English whether it is their first or additional language are able to use English for real-life communicative purposes through

¹ Keynote paper presented at International Conference on Multimedia Language Education, Thirteenth (ROCMELIA 2009), 30-31 Oct 2009, Kaohsiung, Taiwan.

the Internet and various mobile technologies, with motivation as never before. Furthermore, this functional, purposeful usage gives instantaneous feedback to the user. It is much more motivating compared with traditional language learning that typically confines the learning to the classroom where it is difficult to create purposeful, functional engagement. The increasing availability of Internet access now provides a potential key to unlock second/foreign language acquisition through online learning and assessment. This enables learners to better understand their language proficiency and have greater input in and control of their learning. For teachers it can enhance their pedagogical approach, help monitor progress and support collection of data for assessment. Learners may also have opportunities to self-assess which has the potential to reduce test anxiety (Mousavi, 2009) when it comes to taking high stakes tests for education and work such as IELTS and TOEFL.

Thus, there is great pressure on the global citizen of today to become proficient at English, as O'Neill and Gish (2008, p. 19) note “. . . as the opportunity for international travel, international education, transnational qualifications and jobs looms ahead . . . more people are becoming internationally mobile [this] highlights their need for effective skills to communicate . . . in the English language”. So, English can be described as an international language that is needed to communicate and interact for study, for using the many services on the Internet, for work and global social networking. It may be argued that without English language skills and Internet access people today are disempowered and isolated from the world and society because they are unable to participate. The provision of feedback to learners and their teachers on acquisition and improvement in English proficiency besides possession of tools to assess suitability for entry into education courses taught in English or for employment, is therefore a major priority today.

English language tests also need to provide feedback that places test takers at different levels of proficiency to differentiate standards and to make test results' decision making evidence-based. They also, ideally, need to provide feedback in the form of positive washback (Cheng, Watanabe, & Curtis, 2004). This means that tests as far as possible should provide feedback to teachers and learners alike that will assist in planning for improvement “where to from here”. However, this is not always possible because the purpose of the test may constrain provision of sufficiently detailed feedback e.g. high stakes large testing programs. Resourcing issues may also severely impact on the extent to which a test may be able to provide positive washback. For instance, it would be time consuming to score a multiple choice paper and pencil test by hand compared with machine scoring. If fully computerized it may be automatically scored and the results collated and feedback built in at the question or item level to enable learners to learn from their experience. With the latter approach there would be a need to have a large bank of test items and a suitable number of equivalent items thus placing more time and expertise in the design and maintenance of the test compared with the scoring and feedback. Similarly, if tests are to be diagnostic in nature then there needs to be careful design of every question or item to identify the skill demand involved. The test designer needs to be clear about the nature of what has to be tested and develop test specifications accordingly. There are very clear advantages for using online tests such as NETPAW

because of the streamlined management and automatic data gathering and collation. Such an approach reduces resource demands for marking and has the potential to increase validity and reliability because of the automation and control of content, as well as presentation and accessibility.

From a pedagogical perspective it is important to note that many variables may impact upon language learners' acquisition however for the purposes of this paper only those seen as most influential for test development will be considered. These are the classroom pedagogical approach as reflected in the nature of the learning experiences, the learners' opportunities to use the language for real-life purposes, and the approach to testing and assessment. Today's learners and teachers of English have more options than ever before to use learning strategies that provide opportunities for interactive communication and feedback. However, if learners are not involved in interactive communicative learning experiences then they cannot be expected to score highly on tests that are of that format. By contrast, if they are working through a set text book and focusing on the testing of grammar then test performance would be stronger on items that reflected that experience Ingram, Kono, O'Neill and Sasaki (2008). As is well known, unless second/foreign language learners have the opportunity to use the language for real-life communicative purposes their acquisition is slow and often ends in fossilisation. In addition, time spent learning the language is a key variable. Many language learners embark on courses that do not realistically provide enough hours to meet learner expectations. This is in spite of the fact that learners will always need to spend extra time using the language and maintaining their skills. On this basis, being able to experience and gauge one's success as a language learner is essential as is the provision of regular positive feedback if engagement is to be established and maintained. However, it not easy for teachers to develop tests and assessment strategies at the level of sophistication that may be necessary given their time consuming work. As well, development requires high level knowledge and skills in test design and for dealing with the construct in question and the associated issues of validity and reliability.

For provision of the most useful feedback for learning, tests need to be designed to be diagnostic. Diagnostic tests are used to diagnose students' strengths and weaknesses so that programs of learning can be devised which best suit students' needs. Diagnostic testing may range from the use of a proficiency test through use of a test targeting one area such as grammar to the prudent teacher's observation or interview with the student (Hughes, 2003).

As Bachman and Palmer (1996, p. 209) stress when deciding how to score diagnostic language tests "multiple scores and partial credit scoring need to be based on a clear specification of the components of language ability required for correct responses, and the criteria for correctness". They note the importance of relating scoring to the construct that is being measured. They also draw attention to the fact that "neither multiple marks nor partial credit scoring should be used simply to resolve uncertainties about the relative correctness of specific responses, but should be implemented in a principled way, deriving from the design specifications of the test itself" (p. 210). However, they go on to point out that designing for multiple scores or partial credit is more complex because the

criteria for each level of scoring needs to be specific. Similarly, analysis of results becomes more complex and time consuming such that one has to ensure the test and scoring is suitable for purpose. Large testing programs that are well resourced use partial marking effectively and use appropriate statistical analyses like Rasch/Item Response Theory. This has great benefit for education systems and can give rise to useful comparative information for the groups involved and results are reportable at the individual child level, the school, like schools' group, state and national level (e.g. NAPLAN, 2009). Comparative analyses of various groups' performance can be revealing and informative for future curriculum and teaching improvement and also test taking. As happened in the first administration of NAPLAN in 2008 it was realised that states whose students were not experienced in taking such types of tests. These students were said to be lacking in "test wiseness". Clearly for diagnostic purposes at the classroom level and individual student level feedback needs to be user friendly.

In summary, testing of English continues to grow as do the number of English speakers/learners across the world. The need for tests and assessment practices that provide timely, useful feedback to students is crucial for both students and teachers. The need for test results to be linked to a universal standard is also vital to the provision of any feedback as is the ability of a test to provide information that shows students and teachers where improvements are needed. Thus, for a test to be diagnostic it needs to be precise in its identification of the skills under test and have the capacity to add value to the testing exercise as a whole. Bearing the above discussion in mind this research explores the usefulness of NETPAW in contemporary times for administration with international students in Australia.

Approach

NETPAW was explored from three perspectives firstly, from the experience of three English as second language, international student users in Australia, secondly from comparison of this experience with that of an existing diagnostic test of English as a second language (DELTA) (McQueen & Aldous, 1994) and finally through the existing available information on NETPAW (Chuang, 2004; Chen, Chang & Chuang, 2007).

The three students comprised a convenience sample and were of different nationalities and first language background. Two had been in Australia for approximately eight months while the third had many years of experience living and studying in English speaking countries. The DELTA test package was developed through the Australian Council for Education Research (ACER) for year ten and eleven high school students from non-English speaking backgrounds. It contains all the materials to assess listening, reading and writing and the outcomes are presented on a diagnostic map with exemplars. The test development process allows for a testee's performance to be allocated to levels and items/skill demands are specified. The diagnostic nature of NETPAW is explored through comparison with relevant aspects of DELTA and sampling of the two tests skill/item demands. The findings are drawn together through discussion of NETPAW in relation to the usefulness qualities of interactiveness, practicality, authenticity and validity (Bachman & Palmer, 2000) within the limitations of the exploratory nature of the study.

Considerations and discussion

DELTA

For the DELTA listening test testees are required to listen to an audio tape that refers to what should be familiar experiences such as school. First, testees are administered practice items. Then they listen to the tape for the actual test which refers to information in the test book (including choices to tick and some supporting pictorial line drawings). The pace of the speaker tends to be fast. They record their answers in the book and the amount of reading and writing is minimized to ensure the focus is on listening. For listening, the skill underpinning each question is identified as is the level of difficulty (through the initial test development process which used Rasch analysis to arrive at the difficulty levels of the skills under test). The tester manages the time and operates the audiotape player. The total duration of the test is 45 minutes (including 10 minutes orientation). Each question answered correctly is noted by the score on a diagnostic map (by shading the question number if correct). The question numbers (1 to 41) are placed on a vertical continuum moving from easy to harder according to their relative difficulty on the scale. This graphical information portrays the testee's performance by comparing his or her correct answers with the incorrect answers. Because the skill difficulty level is known for each item it is possible to identify the testee's strengths and weaknesses. For instance, a learner may have been successful in listening for the time a train left the station but may have difficulty understanding cause and effect after listening to radio accident report. Similarly, the reading test is contextualized to the learner group and the aim is to make it as communicative as possible. The diagnostic map again identifies whether learners can read to carry out easier tasks such as *find specific information on a time table* and also complete more difficult tasks such as *infer the meaning of a word from the context* or *understand comparison*.

The test materials for writing include stimulus topics that require the writing of two different genres: a science report and the expository genre of a discussion. Writing is scored on the six language features of: *response to the task, register, cohesion, range and complexity of structures, verb use and formation, and vocabulary*. Each feature is scored either one, two or three. These scores represent increasing levels of sophistication developed through the test developmental process. So a testee's scored script identifies the level of performance for each language feature. The package contains scored samples of students' writing to assist testers. Since the scoring criteria are descriptive the reasons why a script belongs to a particular level is also made explicit and a glossary of terms is included. For instance, a narrative that used connectives such as *however, although, in fact, first, secondly, usually, after, before, until, while, eventually, during, consequently* and *therefore* would score more highly than one that confined connectives to *and, then, so* and *also* (O'Neill & Gish, 2008). DELTA (1994, p. 24) also points out that it "assesses first-draft writing composed in a limited time".

DELTA does not test speaking. Obviously as a stand alone package it would have been too difficult to develop a valid and reliable test of speaking with the technology available at the time of its development. It would have had to be in the form of an individualised

face-to-face interview (Ingram & Wylie, 1979/1999), for which the tester would have needed appropriate training and implementation, would have resource implications. Consideration of the test's construct validity shows that it has been developed on the basis of "the Bachman (1990) model of communicative language ability which includes both knowledge about language (grammar and syntax) and the ability to use knowledge in practice" (McQueen & Aldous, 1994, p. 42). It also sets out to differentiate between language used for social interpersonal communication and that required for academic use in keeping with Cummins (1983). This is because "students with facility only in the social domain are known to function poorly in an academic setting" (p. 42). Content validity is also considered and this is argued on the basis of the process of development which drew heavily on the literature in the field. Assessment scales were also researched and the content related to the school materials that the target students were likely to be working on in their school. Since the students were likely to be relatively new arrivals the use of technical terms were not included since it was not expected that this group of students would have them in their vocabulary. Thus, it can be argued that DELTA has substantial validity and authenticity and to some extent practicality for its time and format. With respect to its interactivity it is very much dependent on the learner responding to the spoken or written texts from generally a passive position and so is limited in its ability to satisfy this quality. While, today this test may be seen as out of date technology it remains relevant to its purposes as a diagnostic test. Its developmental procedures with respect to the scales are most appropriate, and the approach to scoring mirrors that of many large scale literacy tests that link to benchmarks and standards today.

NETPAW

NETPAW is a large and ambitious initiative that was supported by the Ministry of Education in Taiwan. It was a key "E-era Manpower Development Project, one of ten projects under Challenge Year 2008—Important Projects for National Development" (Chuang, 2004). It was appreciated that an online test of English language proficiency that used multimedia would be of exceptional benefit for all people in Taiwan and also internationally. Through its establishment of a national testing system it was seen as being able to provide vital information that could be used to improve teaching and learning and so in the longer term improve the English proficiency of the nation. NETPAW was seen as providing a baseline for English proficiency improvement.

NETPAW's specific aim is to promote online English language learning and improve English proficiency. It set out to create an efficient testing system that would be motivating through its ability to provide feedback on students' test results via the computer. It also fulfills an environmental goal to have a paperless test. The first five levels of proficiency are Beginning, Basic, Low-intermediate, Intermediate, and High-intermediate. The higher levels are Advanced and Professional. Language instruction consists mainly of three components: teaching materials, teaching methods, and language acquisition evaluation. As Chuang (2004, p. 6) notes "it is special in that it can provide multimedia and hypermedia, allowing teachers and students to instruct and learn the target language in a non-linear and creative way. It helps promote instruction because of its immediate, international and integrative features". According to Chuang (2004) NETPAW is able to test all four macro skills of listening, speaking, reading and writing and is criterion-referenced and may also be integrated with

online multimedia teaching. While the test can be completed at test centres the approach is flexible in that tests “can be administered automatically anytime and anywhere [and] . . . feedback is given immediately.” (p. 2). NETPAW also sets out to model good practice in English language resources with the results of testing able to feed into evaluating learning programs and provide positive washback to teachers. The use of the web is expected to facilitate and encourage people to use English and also test their proficiency.

The listening test engages the testee in an interactive manner from the start. It allows for the testee to have a practice and also provides a small amount of information to read to help the testee understand the test format and remind the testee what is being tested and that the test is being timed. The screen format is shown in Figure 1. On each occasion the testee listens to an episode of speaking. This is followed by the speaker asking a question to gauge whether the testee understood what was being said. The testee is required to read four possible answers and then choose the correct answer by clicking against a, b, c or d. Thus, there is an element of reading involved. The skill demands of the questions vary and have broad coverage and overall are linked to levels within the framework shown in Appendix A. Compared with the more traditional listening tests like DELTA, administered by a person, NETPAW’s listening test achieves a more natural situation because of the communicative topical nature. The conversation is directed to the testee rather than the testee being in the position of “the observer/listener”. It may be seen as more authentic in the nature of the communication for the testee and the testee remains in control of his or her participation by being able to start, pause and stop the audio as well as restart and listen to the talk again. This allows the testee to skim the choice of answers and also play the talk again to check.



Figure 1: NETPAW format showing media player format

Another advantage is that the testee remains aware of the time as the digital display shows the minutes decreasing. Further, time limits or duration of the test seemed to be reasonable and representative of the speed of comprehending the spoken message in a real conversation/language exchange. But it must be remembered that the testee may listen more

than once and pause at any time, while also needing to listen to the multiple choice possibilities. One could argue that there is an element of integration of skills (listening and reading) but in the absence of being able to have an interactive conversational interview this test is the next best approach.

Figure 2 and Figure 3 (taken from Chuang, 2004, pp. 4 and 6 respectively) show a Basic Level listening task and a High-Intermediate Level listening task. At the Basic Level the testee hears A. "I can see two zebras." B. "I can see two horses." C. "I can see two elephants." S/he then must select A, B or C as the correct answer by clicking with the mouse. This skill demand is to *listen for specific information, matching what is heard to a picture*. At the High-Intermediate Level the testee listens to a conversation between two people:

Mike: "Why do you think God allows evil to exist on Earth?"

Alice: "Well, many people have asked that question, and they have different opinions."

Mike: "Yes, but what do you think?"

The speaker then asks "What are Mike and Alice discussing? The testee needs to read the possible four short answers of two or three words and use the mouse to select their preferred choice. This skill demand is to *listen to a short two-way conversation and identify the topic*. These are in keeping with the skills tested by DELTA test.



Figure 2: NETPAW Basic Level listening task

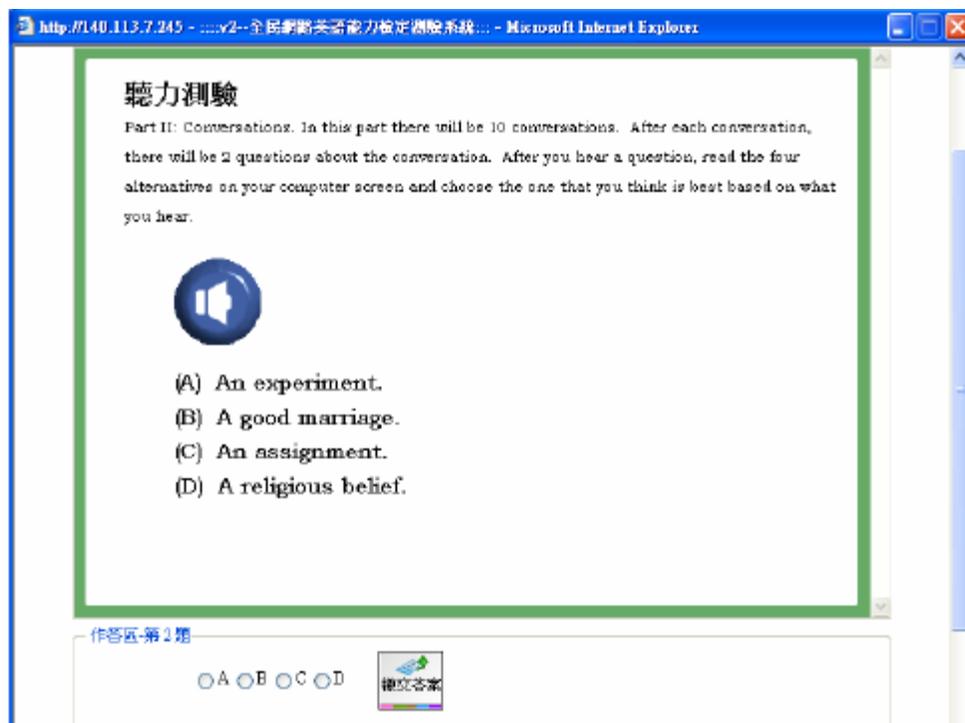


Figure 3: NETPAW High-Intermediate Level listening task

When the NETPAW reading test is considered the approach to content and comprehension questions are comparable with DELTA. Both test are diagnostic in their ability for instance to differentiate between students ability to read and understand shorter and longer written texts that are likely to be familiar in content. They are similar in format except NETPAW is online. Similarly, they differentiate levels of skill difficulty. This is for example, with respect to testees' ability to *understand the main point of a short text*; to *read for explicitly stated, specific information*; *skim for the main idea, understand pronoun reference, use information gained from a written text to infer attitude or opinion and understand cause and effect*.

Testees may only proceed to NETPAW Speaking and Writing tests after passing both Listening and Reading tests. In the testing of Speaking through NETPAW the testee reads a short, simple instruction about what to do. This amounts to the testee starting the media player, listening to the speaker and then recording a response to the speaker's instruction. There is also brief but important written information that explains how the marks will be weighted (pronunciation, intonation and fluency along with grammar and vocabulary). The test questions encourage an extended answer and topics have application to everyday life. They may also involve pictorial information. As with NETPAW, in general, the testee exercises some control of the process within the overall time frame. Although this is not a face-to-face, two-way interactive interview, the testee does have to listen to another person and respond accordingly, therefore adding some element of authenticity.

When testees take the NETPAW test of writing they must first read the stimulus item. The content of the stimulus items vary in topic and in length and specify how many words the testee should write. Topics relate to everyday events so that testees will not be penalised through lack of knowledge. A note is displayed to advise testees of the marking

criteria (content, grammar and spelling). The amount of reading ranges from a single sentence to a discussion text (depending on level and genre being tested). In each case testees have the opportunity to bring their own knowledge and experience to the writing task. The number of words written by the testee are counted and displayed as the writer types. There is a designated place to write. NETPAW adopts a criterion-referenced approach such that test specifications provide predetermined criteria that guide the construction of questions. Writing demands that emerge from a typical short writing task such as “Please briefly describe your favourite person” range from using punctuation appropriately (e.g. capital letters, full stops, placing commas in series); writing grammatical correct sentences, using reference and suitable connectives to achieve cohesion (e.g. subject-verb agreement, backward and forward references, use of ellipsis, noun-pronoun reference, possessive pronouns). In a similar way to DELTA, outcomes may be linked to levels of performance and the nature of the feedback from scoring is diagnostic in nature.

The NETPAW experience

The NETPAW tests of the four macro skills were explored by the three students and the researcher. The test was introduced through an informal examination of the web site and use of the practice materials to allow the students to experience the format and approach. Together they first completed a listening and reading test. Approximately, ten days later the speaking test was explored and then the writing test. Following this the students were asked to think about their experience and respond to the following questions:

1. How did you find the test?
2. What did you like best about it? Can you tell me why?
3. What other types of English tests have you taken?
4. How does NETPAW compare with the tests you have taken before?

Students’ perceptions of NETPAW

The students were pleasantly surprised at the format and ease of use of the test. They viewed it as a valuable resource for learning and practicing English as well as for testing. Initially, they came to the situation with some reservations because of the online nature of the activity and also because of the connotations of “test”. One testee recommended that more information be provided about the test in advance of the introduction (Student A). However, they began to feel more comfortable when they realised that they had some control over time and the topics did not seem onerous. In addition, they were aware that they were not being formally tested. While, there was some concern about time, as there would be with any test, in the case of NETPAW, testees work through the test and operate the media player type format as they go along. At first the testees were involved with reading the test instructions and information on the marking criteria (although these texts comprised short sentences and they could be revisited at any time at a click of the mouse). Similarly, there was a need to become familiar with using the media player style format to respond and self-manage working through the test.

Becoming familiar with the test format was important to the students as this was a new experience within the context of using a computer. The fact that the test instructions allowed them to choose when to start the test was viewed positively as was the way the

format allowed them to quickly change/edit their responses. In addition, the requirement to type or speak responses directly into the system was well liked. This was because of the efficiency. Also it was encouraging in that the testee chose when to respond and whether to listen to a test question again or spend more time re-reading the stimulus text. While being able to see the seconds ticking away might have been a concern at first, as the testees worked through the questions they began to feel more confident about their time management. They gradually overcame the worry of completing as they realized that the time allowed seemed to be appropriate. Nevertheless, there was no time to waste.

As with a conventional test all the information about how to complete the test was available except testees could not see the content of the questions beyond the one they were working out. The need to monitor progress was met by the fact that the number of questions is made clear through the use of a number line. Each unit (a square) represents a question and as they are completed the unit changes colour. The students viewed these aspects of the format as very supportive. They felt responding via the computer was helpful as in the case of the writing test they found it encouraging. This was because they didn't have to worry about mistakes and alterations as in hand written tests. This also applied to speaking since they could playback and re-record their response, but at first they were reluctant to speak into the system. Thus, once familiar with its operation this format was seen as an advantage.

All three students had typically learnt English as a foreign language in their home country and two had taken IELTS, one more recently than the other. They had all taken other tests while learning English and were currently studying in English. While two were new to undergraduate study in Australia the third had been in the country for several years and was studying at the postgraduate level. By the end of the exploration of NETPAW the students were enthusiastic about its features and saw it as a very valuable resource for learning and practicing English as well as for testing. The main advantages were seen as the element of independence and interactivity in that they could control their responses and review and edit them as they wished. Student B who had experience with TOEFL noted:

Traditionally TOEFL has relied on multiple choice, fill in the blank and similar gimmicky ways of testing language skills, because such tests are easy to mark. Unfortunately these methods create problems as test takers focus on the technique of doing these tests rather than on the language skills themselves. But the method used by NETPAW focuses on the interactions among areas of language knowledge and strategic competence or meta-cognitive strategies enables test taker to interpret language that is appropriate for given test situation.

They thought the computer/multimedia approach was an asset to language testing. As a result of this exploration it was concluded that NETPAW was a very worthwhile test and a resource that these students viewed as highly relevant to improving their English language proficiency level as well as assessing it.

Conclusion

This research comprised a small exploratory study of the National English Test in Proficiency for All on the Web (NETPAW) as a diagnostic test for use with English as a second language, tertiary students in Australia. It was limited to a descriptive study that comprised preliminary work to design a larger study of NETPAW's application in English Language Intensive Course for Overseas Students (ELICOS). This study will consider NETPAW from the perspective of both assessing students' English proficiency and collecting feedback to improve students' learning and teacher's pedagogical approach with the use of technology.

Within the limitations of this present research it can be concluded that NETPAW improves on traditional paper and pencil testing, and the use of audiotapes operated by teachers or testers because the online, multimedia format is more engaging and authentic for the testee. The incorporation of multimedia into the test materials is a major strength of NETPAW. This is because it not only increases the scope of genres to be tested, it can also cater for the 21st century multiliteracies environment in which students today are immersed. This highlights the greater authenticity of the NETPAW experience and the limitations of traditional tests of English.

The experiences considered here were at the B2 level of high intermediate and were viewed as substantially interactive. In view of the choice of topics they achieved a valuable level of authenticity within the limitations of testees responding to the speaker. While the "conversation" was not face-to-face, two way interaction the style and format was suggestive of such a context. The format was also seen as meeting the quality of practicality because of the user-friendliness in the testee being able to easily operate the system. NETPAW has a high level of capacity to develop students' English proficiency because its interactive nature requires the use of English for real-life purposes to operate the system and complete the test. Because this builds in speaking it targets the skills that are required first and are the most difficult to learn without communicative practice. It also provides a more integrated use of language which again contributes to its authenticity while carefully minimizing the skills that are not under test. While it may be argued that the tester uses English to complete a pencil and paper test NETPAW transcends the laborious aspect of hand writing and proofreading and editing. It manages to speed up the process by using the mode most commonly used for communication in testees' social worlds today, through an effective electronic solution. It is not surprising then that Chen, Chang and Chuang (2007, p. 38) note "experts predict that in the next few decades, over 50% of the student population will be educated using online learning and technology".

It can be appreciated that the NETPAW test experience can also be a learning experience particularly if users have the time to interact for longer. It can be argued also that as with all tests testees will achieve better when they are familiar with the format. As a diagnostic test NETPAW provides individual reports for each skill. These reports list a testee's English language strengths and weaknesses and include specific remarks that are designed to help the testee improve. They identify the testees' capabilities such as *You can understand the main ideas of complex speech on both concrete and abstract topics*

(listening), *You can read correspondence relating to your fields of interest and easily understand the essential meaning* (reading), *You can produce stretches of language with a fairly even tempo, but you can be hesitant as you search for patterns and expressions* (speaking) and *You can construct a chain of reasoned argument* (writing). In conjunction with the actual experience in taking the test this type of diagnostic feedback constitutes positive washback for testees' personal use and for their teachers. Both the experience and feedback has implications for the teaching and assessment cycle if such technology is available as a pedagogical approach. As a resource it has the much needed ability to engage students in learning English in a more creative and authentic way without the time and resource constraints of traditional classrooms and textbook approaches. It can provide a flexible approach that has the potential to deal with large numbers of students and create a more motivating learning environment.

Although it must be stressed that this study was small scale and exploratory in nature it is concluded that NETPAW goes a long way to meeting its own goal of promoting online English learning and offers a valuable alternative that is user-friendly through its availability online and its interactive assessment experiences. This is supported by the results of a comparison with a typical diagnostic English language test, consideration of test usefulness criteria and the experience of a small number of testees who participated in an informal trial. Overall, NETPAW has potentially multiple uses and international applicability for English language assessment and the teaching of English. This includes its diagnostic application. The fact that the NETPAW levels relate to the Common European Framework of Reference for Languages: Learning, Teaching and Assessment that describes learner proficiency at each level within a hierarchy for European languages is a further advantage. This assists in making links between NETPAW levels and other ESL/EFL scales and assessments so essential to the interpretation and usefulness of the results.

References

- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Bachman, L. F. & Palmer, A. (1996/2000). *Language testing in practice*. Oxford: Oxford University Press.
- Cummins, J. P. (1983). Language proficiency in academic achievement. In J. W. Oller, Jr. (Ed). *Issues in language testing research*. Rowley, Mass.: Newbury House.
- Carroll, B. J. & West, R. (1989). *ESU framework: Performance scales for English language examinations*. Harlow: Longman Group UK Ltd.
- Chen, F. Chang, T. & Chuang, Y. (2007). Using NETPAW to advance senior high school students' English acquisition in the context of the Common European Framework, *World Transactions on Engineering and Technology Education, UICEE*, Vol. 6, No.2. 1-11.

- Cheng, L., Watanabe, Y. J. & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. New York: Lawrence Erlbaum.
- Chuang, Y. (2004). The *National English Test in Proficiency for All on the Web in Taiwan* (NETPAW), retrieved 02 August, 2009, from www.fas.nus.edu.sg/cls/clasic2004/clasicprog_ss4.htm
- Chuang, Y. Chiang, W. & Beasley, R. (2008). The Online CEF Taiwanese Test (TATPAW) Proceedings of the Twelfth International Conference on Multimedia Language Instruction (pp. 16-26). Retrieved July 30, 2009 from <http://www.franklincollege.edu/pwp/RBeasley/Vita.htm>
- Hughes, A. (2003). *Testing for language teachers*. (2nd. ed.). Cambridge: Cambridge University Press.
- Ingram, D. E., Kono, M., O'Neill, S. & Sasaki, M. (2008) *Fostering positive cross-cultural attitudes through language teaching*. Queensland, Teneriffe: Post Pressed. 1-244.
- Ingram, D. E. & Wylie, E. (1979/1999). *The International Second Language Proficiency Ratings (ISLPR)*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University.
- McQueen, J. & Aldous, C. (1994). *Diagnostic English Language Tests (DELTA): Assessment of Listening, Reading and Writing Skills for non-English-speaking background students*. Melbourne: Australian Council for Educational Research (ACER).
- Mousavi, S. A. (2009). Multimedia as a test method facet in oral proficiency tests, *International Journal of Pedagogies and Learning*, Vol. 5, 1. pp. 45-68.
- NAPLAN (2009). National Assessment Program Literacy and Numeracy. Retrieved 01 September, 2009 from http://www.naplan.edu.au/naplan_2009_tests.html

Appendix A NETPAW Proficiency Levels

The NETPAW-CEF Table			
Users	CEF	The CEF Can-do List	NETPAW
Proficient User	C2 Mastery	Can understand with ease virtually everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously very fluently and precisely differentiating finer shades of meaning even in more complex situations.	N/A
	C1 Effective Operational Proficiency	Can understand a wide range of demanding, longer texts and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.	N/A
Independent User	B2 Vantage	Can understand the main ideas of complex text in both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speaker quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various points.	High-Intermediate
	B1 Threshold	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst traveling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes, and ambitions and briefly give reasons and explanations for opinions and plans.	Intermediate
Basic User	A2 Waystage	Can understand sentences and frequently used expressions related to area of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in area of immediate need.	Elementary
	A1 Breakthrough	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.	Basic Beginning

From Chuang, Chiang & Beasley (2008, p. 5)