



## Error deduction and descriptors – A comparison of two methods of translation test assessment

*Barry Turner*

*Royal Melbourne Institute of Technology*

[barry.turner@rmit.edu.au](mailto:barry.turner@rmit.edu.au)

*Miranda Lai*

*Royal Melbourne Institute of Technology*

[miranda.lai@rmit.edu.au](mailto:miranda.lai@rmit.edu.au)

*Neng Huang*

*Nanjing University of Finance and Economics*

[nenghuang@yahoo.com.cn](mailto:nenghuang@yahoo.com.cn)

**Summary:** This paper examines two assessment methodologies used for large-scale translating and interpreting accreditation testing: error analysis/deduction and descriptors. A report by the Royal Melbourne Institute of Technology (RMIT University) (Turner and Ozolins, 2007) showed that the U.K Institute of Linguists and the American Translators Association are among international testing bodies that have moved or are moving towards using descriptors or combining negative marking and descriptors. This paper explores whether the Australian National Accreditation Authority for Translators and Interpreters (NAATI) might be able to move to a descriptor approach to assessment without risk to the reliability or accountability of its public examination system. The NAATI assessment system is used as a benchmark to compare it with assessment outcomes using the descriptor-based translation component of the U.K Institute of Linguists Diploma of Public Service Interpreting (DPSI). The most significant finding of the research is that there was a high correlation between assessment outcomes in the two assessment systems, indicating that a descriptor system might be as reliable and accountable as the current NAATI system.

**Keywords:** translation; translation examination; negative deduction; error analysis; descriptor; marking; accreditation

### 1. NAATI

The Australian National Accreditation Authority for Translators and Interpreters (NAATI) is a national standards body owned by the Commonwealth, State and Territory Governments of Australia. Since it was established in 1977 it has since run a large accreditation testing program and has approved university and TAFE (Technical and Further Education) courses that lead to NAATI accreditation. RMIT University, like a number of other universities and TAFE providers in Australia, conducts NAATI approved translating and interpreting programs and testing is carried out in accordance with the same guidelines that NAATI applies in its public examination system.

The translation component of the current NAATI Manual (2005) provides a detailed error analysis/deduction system and standardised symbols to identify and categorise errors. On the other hand, the interpreting component provides much less guidance to examiners (Turner and Ozolins, 2007, p.31).

## **2. 2007 RMIT research**

In 2007, NAATI funded RMIT to conduct a research project on the standards of linguistic competence among NAATI accredited translators and interpreters (Turner and Ozolins, 2007). In addition to data collected within Australia, information was gathered on twenty-four translating and interpreting accreditation bodies around the world. The researchers found that accreditation systems in use included:

- error analysis/deduction systems (like NAATI's system for assessing translation tests);
- a combination of criterion-referencing (the use of scales of descriptors to describe performance in tests) and error analysis/deduction; and
- criterion referencing (descriptors) with no system of error analysis/deduction.

One of the findings of this research was that a large number of stakeholders in the translating and interpreting industry considered that test candidates who had gained NAATI accreditation had adequate language proficiency (Turner and Ozolins, 2007, p.9) and that the NAATI assessment system was well received by NAATI examiners (Turner and Ozolins, 2007, pp.27-28). The report drew attention to the recommendation made by an earlier report commissioned by NAATI for more 'holistic' methods of assessing translation and interpreting tests be developed (Turner and Ozolins, 2007, p.102). The research therefore made special note of the adoption of descriptors in the UK and Norwegian translating and interpreting accreditation examination systems and the American Translators Association's move towards using descriptors. One of the report's recommendations was that NAATI consider adopting descriptors in its translating and interpreting examinations instead of the current error analysis/deduction method (Turner and Ozolins, 2007, p.103).

## **3. Rationale for this research**

RMIT's research team set out to explore whether NAATI might be able to move to a descriptor approach to assessment without risk to the reliability or accountability of its public examination system.

The few internationally published systems of assessing interpreting examinations (for example the DPSI and Norwegian systems) use descriptors or holistic approaches. Because of this practice and the relative lack of guidance in the interpreting component of the NAATI Examiners' Manual (Turner and Ozolins, 2007, pp.31-32), we concluded that NAATI could move to a descriptor system for assessing its interpreting tests with little, if any, risk to reliability or accountability.

On the other hand, the translation component in the NAATI Examiners' Manual provides comprehensive guidance in categorising and deducting marks for errors (Turner and Ozolins, 2007, p.45) and is supported by examiners (Turner and Ozolins, 2007, pp.27-28). Therefore, we decided to concentrate on the assessment of translation tests to try to determine if it would be feasible for NAATI and NAATI approved training institutions to adopt a descriptor-based assessment system that would be as reliable and accountable as the current system.

We came to the view that researchers in Australia are in an unusual and possibly unique position to conduct similar research. Because educational institutions that run NAATI approved courses must assess students in the same way that candidates in the NAATI public examination system are assessed (in accordance with the NAATI Examiners' Manual), there is a uniformity of approach between assessment standards and methods under the national public examination system and in NAATI approved training courses. We decided to compare assessment outcomes at RMIT between the use of descriptors and the current NAATI error analysis/deduction system.

#### **4. Research framework**

The 2007 RMIT Research drew particular attention to the descriptor system used by the U.K Institute of Linguists in the translation component of its DPSI tests (Turner and Ozolins, 2007, p.103). These descriptors are publicly available on the internet (<http://www.iol.org.uk/qualifications/DPSI/DPSIHandbook.pdf>). Because they were readily available and had been cited in the 2007 research, we decided to use them for this project.

We were aware that the internet version of the DPSI descriptors might have been truncated and our examiners would not be properly trained in their use. Nevertheless, we considered the DPSI descriptors sufficiently detailed for our purposes and all of the examiners we chose were highly experienced translation teachers and examiners who, after a brief training session, would be able to interpret and use the descriptors to an adequate level.

Our aims were to:

- identify the extent to which results obtained from DPSI descriptors would correlate with the NAATI error analysis/deduction system; and by doing so,
- compare the reliability of the DPSI assessment system with NAATI's.

Our intention was to explore whether the reliability and accountability of NAATI's translation test assessment system might suffer if NAATI and NAATI approved training institutions were to move to a descriptor-based assessment system.

For the purposes of the experiment, we decided to use the NAATI system as our benchmark, and we used the final translation accreditation exams (NAATI Professional Translator level) of students in four language streams at RMIT in November 2008. The exam papers were marked in three stages:

- Stage one: using DPSI descriptors only with examiners being aware of the identities of students;
- Stage two: using the NAATI error analysis deduction system (the normal method of assessing accreditation students at RMIT) with examiners being aware of the identities of students; and
- Stage three: using DPSI descriptors with identities of students concealed from examiners, plus NAATI error annotation symbols.

#### **5. Research findings**

We found that there was a high degree of correlation between the NAATI and DPSI assessment systems. This suggests that NAATI might be able to

move towards a descriptor-based system without sacrificing reliability of assessment.

We also found that correlations between the two assessment systems were somewhat lower when examiners had to operate without knowledge of the identities of the students being assessed in stage three of the experiment. This perhaps indicates that examiners in a university setting such as RMIT might be influenced by knowledge of the identities of the students they are assessing.

## 6. Methods of assessing NAATI translation examinations

NAATI translation test passages at the Professional Translator level are approximately 250 words (or equivalent in English) in length. Candidates must select two passages out of three provided and each candidate is independently marked by two examiners who do not have access to each other's marking during the process. There is a cut-off average percentage of 70% for the two translation passages below which test candidates are deemed to have failed.

The translation section of the NAATI Examiners' Manual gives examples of errors of different severity, and guidance on the points (from half a mark to five points or more) that should be deducted in each case. It categorises errors in terms of accuracy, quality of language and technique.

The Manual provides a system of symbols to indicate whether mistakes affect accuracy or quality of language. For example, words deemed inaccurate are enclosed in a rectangle; errors affecting accuracy are annotated in solid lines and those relating to quality of language are in dotted lines.

The following example illustrates how marks deducted are annotated under the NAATI system. The title of a source text is shown in the first row and the back translation (into English) of a candidate's translation is shown in the second row. The marks deducted for the translation, in order of the appearance of errors in the back-translated text, are shown to the left of the candidate's response in the second row.

### Example of NAATI marking for translation errors:

<b>Source Text</b>	<i>Pollution Advocacy: Demanding the Public's Right to a Clean Environment</i>
2 1 ½ ½	<i>Pollution <span style="border: 1px solid black;">Lawyers</span> <span style="border: 1px solid black;">Suing for</span> the Public's Right <span style="border: 1px dotted black;">for</span> a Clean ( <del>Living</del> ) Environment</i>

If a similar number of marks per line of text were to be deducted as in the example above, it would not take long for a candidate to be penalised sufficiently to fail to achieve accreditation standard. Accordingly, pass rates for NAATI-conducted Professional Translator tests are generally low with only 18.02% achieving accreditation at the Translator (into languages other than English) level. Interestingly, despite the relative lack of guidance in the interpreting section, pass rates are similar with only 15.22% of interpreting test candidates passing in NAATI's 2005-06 annual testing program (Turner and Ozolins, 2007, p.18). These rates appear to be similar to those of similar overseas testing organisations (Stejskal, 2003, pp.30, 37).

## 7. Descriptors in language proficiency and translation assessment

The use of descriptors in the language field has been pioneered by organisations that assess general language competence. One of the best known descriptor-based systems for assessing language competence is the International English Language Testing System (IELTS). The introduction of holistic systems such as the IELTS has resulted in greater sophistication in language assessment. Unlike previous error analysis or intuitive and non-standardised means of assessing language, IELTS and similar descriptors describe the way test candidates use language beyond the level of a particular word or grammatical category which can be assessed as right or wrong. Importantly, they describe what a candidate *can* do as well as what they *cannot* do.

Accuracy is probably the paramount consideration in assessing translation and interpreting tests and a descriptor system for translating and interpreting tests must ensure that the outcome is a faithful rendition of the original source text. Descriptors that fail to take sufficiently into account these particular aspects of translation and the consequent assessment of translations could result in examiners failing to take sufficient note of errors in the target text.

Assessors who used descriptors in language proficiency assessment, such as IELTS, undergo extensive training to enable them to interpret and apply them and to try to standardise assessment as much as possible between examiners. If NAATI were to adopt descriptors, it would have to consider the need to train all the members of its almost sixty language panels, some of which are drawn from societies where translating and interpreting is relatively undeveloped as an academic discipline (Turner and Ozolins, 2007, p.29). On the other hand, RMIT runs programs in fewer than ten languages each year and unlike most NAATI examiners, RMIT staff members teach translating and/or interpreting throughout the year and receive ongoing in-service training in teaching and assessment. This facilitated our use of RMIT examiners in this research project.

## **8. The diploma in public service interpreting assessment system**

The DPSI divides its assessment criteria for translation tasks into “accuracy/appropriateness of translated text”, “cohesion, coherence and genre conventions”, and “effectiveness of communication” (UK Institute of Linguists, 2007).

DPSI translation candidates are assessed according to these categories and within four bands of descriptors (A, B, C and D). Like the IELTS system for assessing language competence, these bands describe what test candidates can and cannot do. For example, a candidate who performed exceptionally well in “accuracy/appropriateness of translated text” in a DPSI written translation task would fulfil the following criteria (Band A):

The candidate

- conveys information with complete accuracy
- conveys all information without omissions, additions or distortions
- shows excellent use of vocabulary throughout
- uses excellent grammatical/syntactical constructions
- displays faultless spelling, accentuation/use of diacritics, faultless punctuation
- has excellent knowledge of specialised terminology with minimum paraphrasing

In addition to these criteria, the DPSI system also uses numerical marks to differentiate between levels of proficiency within the same band. Each band level is given a mark range and the final overall grade for each exam component is reached by adding up the marks for each assessment category (see below):

#### Numerical Marks for Each Band in Each Assessment Category

Band D	1- 3
Band C	4- 6
Band B	7- 9
Band A	9-12

#### Overall Grade for Each Exam Component

Fail	Fewer than 12 marks overall or fewer than 4 marks in any one category
Pass	A minimum of 12 marks overall with no fewer than 4 marks per category
Merit	A minimum of 21 marks overall with no fewer than 4 marks per category
Distinction	A minimum of 30 marks

### 9. Holistic guidance

An advantage of descriptors, such as those used in the DPSI, is that they offer more holistic guidance to examiners than is available in the NAATI error analysis/deduction system. Two of the researchers for this paper (Turner and Lai) have been involved in supervising the assessment of accreditation tests at the NAATI Professional Translator level in an educational context for some years. They have sometimes felt the need for such descriptors at RMIT, where they have concluded that some examiners become too fixated on classifying errors and deducting points in accordance with the NAATI Examiners' Manual, to the extent that it impairs their ability to arrive at a holistic appreciation of the translation they are assessing. In some cases, a tally of the points deducted has led to students being assessed at accreditation level, whereas it is obvious to a reader who takes a more holistic view that the test is in fact below the required level. On the other hand, there have also been cases where too many points appear to have been deducted in a mechanistic way and the student was assessed to have failed while the translation was actually at the accreditation standard.

### 10. Detailed and consistent feedback to test students

At RMIT, percentage scores are supported by notations on the students' test papers and comments by examiners, but these are sometimes not adequately or clearly expressed. Moreover, the use of percentage scores in this way implies a mechanistic process, rather than a qualitative one. Students sometimes attempt to 'haggle' with examiners over points to be deducted for particular errors, rather than looking at the quality of their translation at a holistic level.

Turner and Lai have sometimes felt that if results were expressed in terms of a level of proficiency within a graded system of descriptors, students might focus more on their overall performance than upon a percentage score and that if examiners had descriptors to refer to, they would be able to give more

appropriate and comprehensible reasons for their decisions. In particular, they would be able to provide standardised and meaningful comments regarding performance.

## **11. Preliminary research at RMIT**

In early 2008, two staff members at RMIT experimented with using the UK DPSI descriptor system. It was interesting that both examiners found it useful to employ a standardised set of symbols to identify and categorise errors in addition to the DPSI descriptors because they were concerned that they might not be able to examine the renderings of test candidates in detail unless they identified different classes and types of errors in this way. Of course, both were no doubt influenced by prior exposure to the NAATI system of error analysis/deduction.

## **12. A rigorous descriptor system**

Following this preliminary research Turner and Lai tentatively concluded that due to the particular nature of assessing translation and interpreting tests described in Section 3 (the need to ensure that the outcome is a faithful rendition of the original source text and the requirement for examiners to constantly refer back to the source text), an optimum system for assessing translation tests might incorporate both holistic guidance and a rigorous approach to identifying and categorizing errors.

With these considerations in mind, the research outlined in Section 1.4 was embarked upon in late 2008.

## **13. A comparison of the NAATI and DPSI systems – NAATI professional translator accreditation examinations**

This research involved the use of a total of thirty-six NAATI Translator final Accreditation Examination papers in November 2008, from the following language streams at RMIT:

- English into Chinese (seventeen papers)
- English into Japanese (five papers)
- Arabic into English (nine papers)
- Italian into English (five papers)

Ethics clearance from RMIT University had been obtained before the commencement of the research. Students from these language groups were informed of the research and were advised that only the NAATI assessment system would be used to determine their accreditation results and that their participation of the research was completely voluntary. The papers of those who did not give written consent were excluded from this research.

The translation examination papers used in this research followed normal NAATI examination marking guidelines. Students in the translation examinations had selected two out of three passages of approximately 250 words in length, with each passage accorded fifty points (a total score of

100). Three stages of experimentation outlined in Section 4 were designed for the purposes of this research.

We did not provide our examiners with our estimation of the cut-off point in the DPSI descriptors that we felt was equivalent to the bottom threshold of the NAATI accreditation level. We felt that had we done so, our examiners, all of whom were steeped in the NAATI system, might seek to orient their assessment towards the NAATI system instead of using the descriptors in their own right.

A focus group was held in January 2009 to collect feedback from the examiners and is reported in Section 13.5 of this paper.

### **13.1 Stages of the experiment**

Stage One (referred to as DPSI1 in statistical tables): examiners marked the papers using the DPSI descriptors only. They were aware of the identities of the students (hereafter referred to as ‘non-blind’ marking).

Stage Two (NAATI in statistical tables): immediately after stage one, examiners used the NAATI assessment system in accordance with the normal practice for accreditation tests. They were aware of the identities of the students.

Stage Three (DPSI2 in statistical tables): two weeks after completing stage two, examiners assessed the same translation passages using both the DPSI descriptors and the NAATI error annotation symbols. The identities of students were not provided (hereafter referred to as ‘blind’ marking) and student translations had been typed and printed (including student errors) so that their hand writing would not be identifiable.

The Statistical Package for Social Sciences (SPSS) system was used to analyse the marking results from the three stages of experiments and the outcomes are presented in a series of tables in the following sections of this paper.

The statistical tables in the following sections:

- compare NAATI accreditations and DPSI descriptor outcomes and attempt to determine a DPSI cut-off threshold commensurate to NAATI accreditation level; and
- examine the correlations between stages one and two, as well as between stages three and two.

### **13.2 Comparison of NAATI accreditations with DPSI descriptor marking outcomes**

We began by examining the distribution of DPSI Descriptor outcomes compared with results achieved under the NAATI error analysis/deduction system. The details of the comparison are presented in Table 1 below.

The table shows that:

- Of the thirty-six sampled papers, nineteen received NAATI accreditation scores of 70% or above, while seventeen did not achieve the NAATI accreditation level.
- Of the seventeen papers from all languages that did not achieve NAATI accreditation, the majority were awarded DPSI Pass or lower marks with rare exceptions (one under non-blind DPSI1 and two under blind DPSI2 who were awarded DPSI Merit marks).
- Of the nineteen papers from all languages that achieved NAATI accreditation, the majority achieved DPSI Merit mark or above, with a small number (three under non-blind DPSI1 and four under DPSI2) scoring DPSI Pass marks.



**Table 1: Cross tabulation of the two DPSI applications by language and NAATI accreditation result**

Language	DPSI Outcome	No NAATI Accreditation (<70%)		NAATI Accreditation (>70%)		Total	
		DPSI1	DPSI2	DPSI1	DPSI2	DPSI1	DPSI2
English into Chinese	Fail	1	0	0	0	1	0
	Pass	3	4	3	2	6	6
	Merit	0	0	5	11	5	11
	Distinction	0	0	5	0	5	0
	<b>Total</b>	4	4	13	13	17	17
English into Japanese	Fail	2	0	0	0	2	0
	Pass	1	2	0	1	1	3
	Merit	0	1	1	1	1	2
	Distinction	0	0	1	0	1	0
	<b>Total</b>	3	3	2	2	5	5
Arabic into English	Fail	5	4	0	0	5	4
	Pass	1	1	0	1	1	2
	Merit	0	1	2	1	2	2
	Distinction	0	0	1	1	1	1
	<b>Total</b>	6	6	3	3	9	9
Italian into English	Fail	1	0	0	0	1	0
	Pass	2	4	0	0	2	4
	Merit	1	0	1	1	2	1
	Distinction	0	0	0	0	0	0
	<b>Total</b>	4	4	1	1	5	5
All Languages	Fail	9	4	0	0	9	4
	Pass	7	11	3	4	10	15
	Merit	1	2	9	14	10	16
	Distinction	0	0	7	1	7	1
	<b>Total</b>	17	17	19	19	36	36

Table 2 below summarises the basic statistics (mean, median, sample size and standard deviation) for both DPSI applications under the NAATI-accredited and NAATI-non-accredited categories.

**Table 2: Descriptive statistics - Both DPSI applications by NAATI accreditation**

		DPSI1	DPSI2
<b>NAATI Non-Accredited</b>	Mean	11.68	14.46
	Median	11.50	14.00
	Std. Deviation	4.10	4.88
	Sample Size	17	17
	<b>NAATI Accredited</b>	Mean	25.18
Median		25.50	22.50
Std. Deviation		6.54	5.22
Sample Size		19	19
<b>Total</b>			<b>DPSI1</b>
	Mean	18.81	18.74
	Median	18.00	17.00
	Std. Deviation	8.75	6.47
	Sample Size	36	36

The table shows that:

- There was no significant statistical difference between the overall average DPSI scores for all thirty-six sampled papers in the non-blind DPSI1 marking (average score 18.81) and the blind DPSI2 marking (average score 18.74);
- For those who failed NAATI accreditation, the average of non-blind DPSI1 score (11.68) is lower than the average blind DPSI2 score (14.46). The former is slightly below the DPSI Pass grade (refer to Section 5) and the latter is slightly above the DPSI Pass grade.
- For those who achieved NAATI accreditation, the average non-blind DPSI1 score (25.18) is higher than the blind DPSI2 score (22.58). Both scores amounted to the DPSI Merit grade (refer Section 5).
- Accordingly, the overall variation within the non-blind DPSI1 scores was greater (standard deviation 8.75) than within the blind DPSI2 scores (6.47). This suggests that in the non-blind application of the DPSI marking, students known by the examiners to have performed well in the past might have been marked up while students known to have performed to a lower standard might have been marked down.

### **13.3 Aligning the NAATI accreditation standard with the DPSI scoring system**

Based on the data from Tables 1 and 2, we then sought to see if we could align the two systems so that we could determine the cut-off point in the translation component of the DPSI system that aligned most closely with NAATI's 70% benchmark.

Table 1 ('All Languages') shows that the majority of students in all languages who achieved NAATI accreditation were awarded DPSI Merit or above - sixteen (84%) using DPSI1 non-blind marking and fifteen (79%) using DPSI2 blind marking (out of nineteen).

In Table 2, the mean DPSI scores for the NAATI-accredited category under the two DPSI applications both fall in the low- to mid-twenties (DPSI Merit grade).

Apart from the Italian group, a small number of students in each of the other three language groups achieved NAATI accreditation levels but fell into the DPSI Pass grade category - three (23%) of students awarded NAATI accreditation in the Chinese group under DPSI1, one (50%) from the Japanese group using DPSI2, and one (33%) from the Arabic group using DPSI2. Upon investigation, it was found that the teachers regarded these students as being at 'borderline accreditation level' based on their overall performance throughout their studies.

The overwhelming concentration of Merit marks (under DPSI1 or DPSI2) awarded to those who achieved NAATI accreditation and the fact that most students who failed NAATI accreditation were awarded DPSI Pass or lower grades indicate that an alignment of NAATI 70% accreditation level with the Merit grade of the DPSI scale is reasonably accurate.

This suggests that the DPSI descriptors, even with our limitations in applying them (Section 4), were capable of producing similar results to the NAATI assessment system. The outcome is particularly interesting given that our examiners were not told of our estimation of the bottom threshold of the equivalence of NAATI accreditation level and they could not have oriented their marking towards it. Also of interest is that examiners did not equate the DPSI Distinction level with the RMIT Distinction grade (70% to 80%) which is the cut-off score for achieving NAATI accreditation. We therefore concluded that our statistical method (of identifying the Merit level as being

equivalent to the lower reaches of NAATI's accreditation standard) was sufficiently accurate for the purposes of this paper.

### 13.4 Correlations between NAATI and DPSI assessments

The second step in the research was to determine how closely the results achieved by using the two DPSI applications correlated with those arrived at under the NAATI system, using NAATI assessment results as a benchmark.

The correlations of all results of the thirty-six sampled papers from the three stages of the research (Section 9.1) are shown in Table 3.

**Table 3: Correlations: DPSI and NAATI results, by language**

<i>Language</i>		<i>NAATI-DPSI1</i>	<i>NAATI-DPSI2</i>
<b>English into Chinese</b>	Pearson		
	Correlation	0.898**	0.862**
	Sig. (2-tailed)	0.000	0.000
	Sample Size	17	17
<b>English into Japanese</b>	Pearson		
	Correlation	0.963**	0.522
	Sig. (2-tailed)	0.008	0.366
	Sample Size	5	5
<b>Arabic into English</b>	Pearson		
	Correlation	0.979**	0.779*
	Sig. (2-tailed)	0.000	0.013
	Sample Size	9	9
<b>Italian into English</b>	Pearson		
	Correlation	0.882*	0.942*
	Sig. (2-tailed)	0.048	0.017
	Sample Size	5	5
<b>All Languages</b>	Pearson		
	Correlation	0.827**	0.735**
	Sig. (2-tailed)	0.000	0.000
	Sample Size	36	36

\*\* Correlation is significant at the 0.01 level (2-tailed)

\* Correlation is significant at the 0.05 level (2-tailed)

The table shows:

- There is a strong correlation in all language groups between the NAATI and non-blind DPSI1 marking, with Italian into English having the weakest degree of correlation (0.882) and Arabic into English results having the strongest degree of correlation (0.979).
- The degree of correlation is weaker, but still significant, between the NAATI and blind DPSI2 marking. This weaker relationship is particularly apparent in the English into Japanese and the Arabic into English groups. The degree of correlation for the Japanese group (0.522) is a correlation not significant at the 0.01 level or 0.05 level. The degree of correlation for the Arabic group (0.779) is significant only at the 0.05 level.
- That the small sample sizes of some language groups (Italian into English, English into Japanese) limit the statistical significance of the relationships shown.

### **13.5 Focus group of examiners**

Four examiners took part in the research - one for each language group. All of the examiners felt that there should have been more training provided in order for them to confidently mark under the DPSI system. Throughout their marking, they had been concerned whether their DPSI marking reflected consistency with their marking under the NAATI system.

When examiners were told the correlations between their NAATI marking and the two DPSI applications were high (with Japanese group at a lower degree 0.522 under DPSI2), they all seemed relieved and somewhat surprised. When asked if they could still recognise the identities of students in the blind DPSI2 marking, most said they could recognise some but not all. One examiner commented: "I found the DPSI2 marking more difficult because of the need to employ the NAATI marking symbols. It was as if I was doing it through the NAATI prism all over again."

At the focus group it was discovered that one examiner (Arabic into English) had not used the NAATI error annotation symbols in the blind DPSI2 marking. The other examiners agreed that the use of NAATI marking symbols in the blind DPSI2 marking helped them to "visualise" the mistakes the candidate made, instead of relying on a "mental picture" acquired under the non-blind DPSI1 marking (without employing the NAATI marking symbols).

When the identities of those students who achieved NAATI accreditation but only scored Pass marks were disclosed to the Arabic, Japanese and Mandarin examiners, they largely agreed that they considered all of them to be at "borderline accreditation" level in their classes and although they had achieved DPSI Pass marks, the descriptors to which their performance accorded were what they might expect of students at the borderline NAATI accreditation standard.

When asked which system they preferred, all the examiners expressed appreciation of the NAATI system. They said this was particularly the case in an educational context, where teachers are obliged to give feedback and justify the marks they award to students. When asked if they would prefer the national exam system to change to a descriptor based one, they all expressed various levels of reservation.

## **14. Conclusions**

In view of the limitations described at the outset of this paper (Section 4), we were somewhat surprised that our experiment showed significant correlations between both the non-blind DPSI1 and blind DPSI2 marking and the NAATI benchmark. An interesting aspect of this outcome is that examiners reached it without being told of an estimated point in the DPSI grade that would accord with the bottom threshold of the NAATI accreditation level.

We believe that in a university or TAFE context, such as RMIT, a descriptor system would probably be able to provide students with more complete and holistic descriptions of their proficiency through a graded series of descriptors of achievement while encouraging examiners to look more holistically at student/applicant performance without sacrificing the reliability and accountability of the current error analysis/deduction system.

An important point to note is, had we used less experienced examiners in this research, we might not have achieved such good correlations. If NAATI were to consider replacing its current system with a descriptor system for its public examination system, it would have to take into account the feasibility

of training examiners in its almost sixty language panels and invest appropriate resources into developing an optimum system for its needs.

It was confirmed in the focus group that the NAATI assessment system enjoys wide support from our examiners. However, they were divided over the merits of using detailed NAATI system of marking symbols for annotating and noting errors of various types in conjunction with descriptors. If NAATI is to consider moving towards descriptors it might be useful for it to conduct further research into the merits or otherwise of incorporating such an aide for examiners.

An unexpected but nevertheless significant result from the research was the higher correlations between non-blind DPSII and NAATI marking in all cases except for the Italian into English group. This phenomenon suggests that examiners at RMIT might be influenced in their marking by their knowledge of the identities of students.

A tendency on the part of some RMIT examiners to allow prior knowledge of students to influence their deduction of marks could be construed in a positive light: that staff members use their knowledge of students gained over the time of tuition to try to award an appropriate mark in accreditation examinations. And conversely, staff at RMIT might give lower marks in accreditation tests to students who have performed below the accreditation standard in translation classes.

## References

- Stejskal, J. (2003). *International certification study*. American Translators Association.
- Turner, B and Ozolins, U. (2007). *The standards of linguistic competence in English and LOTE among NAATI accredited interpreters and translators*. RMIT University, Melbourne. (Available at [www.mit.edu/gsssp/ti](http://www.mit.edu/gsssp/ti).)
- UK Institute of Linguistics. *The Diploma of public service interpreting: Assessment criterion statements*. Retrieved March 2007 from [www.iol.org.uk](http://www.iol.org.uk)