

## **A longitudinal and comparative study: the attainment of language proficiency**

**Annie Brown, Kathryn Hill and Noriko Iwashita  
Language Testing Research Centre  
The University of Melbourne**

### **Abstract**

The paper reports preliminary findings for the first stage of a longitudinal study of comparative attainment in four foreign languages which are widely studied in Australian schools. The study used tests of reading, listening and speaking and writing proficiency designed to be equivalent across languages. As well as examining overall attainment for the four languages, the influence of variables such as years of instruction on performance outcomes was also considered. Findings show significant differences in levels of attainment according to language, with students of Japanese performing at consistently lower levels across all four skills. Variables such as gender, home language use and years of study were found to have a significant effect on test performance, but the size of these effects varied across languages and skills. The paper concludes with a discussion of the policy implications of these preliminary findings for school program provision and for the practice of describing school-based foreign language learning achievement in generic rather than language specific terms.

### **1. Introduction**

This paper presents the findings of the first stage of a DETYA-funded project designed to provide school systems with answers to fundamental languages other than English (LOTE) -related questions. These include the length of time needed to learn a language at school, whether some languages are 'harder' to learn than others, the levels of language proficiency which students can be expected to achieve in school language programs, and how much actual language students will know at the end of a school program in these languages.

The study, *A Longitudinal and Comparative Study: The Attainment of Language Proficiency*<sup>1</sup>, was established as a nation-wide long-term investigation of the issues of language proficiency, with the assumption that the same learners would be followed over the period of their secondary schooling to ascertain long-term language gains. Stage 1 of the project, completed in 1999, was designed to provide a 'snapshot' of performance across the four languages—French, Indonesian, Italian and Japanese<sup>2</sup>—by Year 8 learners.

A number of issues were investigated in the study. Of particular concern was relative attainment across different languages, given the range of languages taught in Australian schools. Whilst there is much anecdotal evidence of different levels of attainment across languages, there are few empirical studies. It was anticipated that the findings would contribute to the debate on the appropriateness of describing school-based foreign language learning achievement in language specific rather than generic terms. A second issue of interest was the relationship of achievement in the four languages to the amount of time spent studying the language. This is also of particular concern given the mixed findings to date about the value of an early start to language study.

The performance of boys and girls was also compared on the different languages and skills, as was the impact of a home background in the language on achievement. The latter issue was investigated in relation to Italian only, as this had the largest number of 'background' learners<sup>3</sup>.

---

<sup>1</sup> Acknowledgements are due to Catherine Elder, Jennifer Makin and Sally O'Hagan, who had significant involvement in this project, and to Lis Grove and other colleagues within the LTRC. The authors would also like to thank Brian Lynch of the University of Melbourne for his helpful statistical advice, and Tim McNamara for feedback on an earlier draft of this paper.

<sup>2</sup> French was chosen because it is already well researched elsewhere. The inclusion of Italian allowed consideration of the native/background speaker issue in second language learning. Indonesian and Japanese, each have with relatively few background speakers and provide a character and a non-character based Asian language.

<sup>3</sup> The term "background speaker" refers to the learner with home exposure to the LOTE. A "Non-background speaker" is a learner with no such exposure to the target language (Elder 1997: 1).

### 1.1. Comparative language difficulty

Within the last few years there has been a significant shift within the LOTE curriculum in a number of contexts (e.g. the Curriculum and Standards Framework II used in Victorian schools), from a generic LOTE curriculum to language specific varieties. One view which supports this shift is known as the 'language distance' theory. Proponents of this theory argue that the more similar a language is to the learners' first language, the easier it will be to learn. Odlin (1989: 153), for example, cites the Foreign Service Institute (FSI) language teaching course length figures to claim that the degree of similarity between languages is a major "determinant of the amount of time students will need to become proficient in a language". Kirkpatrick (1995:8) also uses 'language distance' arguments to explain why he believes Chinese, Japanese and Korean are more difficult than others: they use different scripts to English, they do not use a phonetic system like English, they are not cognate with English, and Chinese is a tonal language. On this basis he claims that first language English speakers would take four times as long to attain basic proficiency in these languages as they would to learn the 'easier' languages (e.g. French). In the same vein Gibbons (1994:16) recommended the introduction of Indonesian, rather than Chinese or Japanese, to the school surveyed in his study because he believed that there was a greater likelihood of students reaching a high level of proficiency in that language.

However, these findings have been criticised for not being based on appropriate evidence. Davies and Elder (1997: 94), for example, question the extent to which the Foreign Service Institute (FSI) figures can be used as evidence for the 'language distance' theory: "why Arabic, Chinese, Japanese, Korean and Polish should be lumped together ... [is] unclear except on narrowly institutional grounds". They suggest that the relative difficulty assumed by the FSI courses may be based more on perceptions than actual evidence. Kirkpatrick (1995) has also been accused of lacking real evidence for his assertions (e.g., Willis, 1995; Gao, 1996).

By contrast, Ringbom's research (1987) does actually provide evidence in support of the 'language distance' theory, finding clear differences in the performance of Finnish (more 'distant') and Swedish-speaking (less 'distant') Finns on formal English language tests. He also found that the effect of 'language distance' is stronger at early stages of learning and at lower levels of proficiency.

### **1.2. The effect of the number of years of study on attainment in language learning**

Whilst it would generally be assumed that learning a language for longer would lead to higher proficiency, research indicates that time alone is insufficient as a predictor of relative second language proficiency (e.g. Harley, 1986) and that condensing the period of time spent on LOTE is better than spreading the same amount of tuition over a longer period (Genessee, 1995). Yet in Australia, as elsewhere, there has been a push to extend the period of school LOTE instruction by commencing it in the primary school. This policy is based on the view that an early start to LOTE learning (in primary school) leads to higher levels of ultimate attainment than would be achieved by commencing LOTE study in the secondary school. However, concerns have been voiced that this is not necessarily the case, and research into the value of starting the study of a LOTE in primary school has produced mixed results. A recent European Union-wide review of research in this area found that commencing LOTE study at primary level did not appear to make a substantial difference to language attainment at secondary school (Blondin et al. 1998). Even where an advantage for an early start has been found, however, it does not seem to persist over time. For example, Genelot (1996) found that although the 'early start' students in France were slightly ahead after the first year of secondary, this advantage had disappeared by the end of secondary school. Likewise, in Australia, two studies comparing the performance of beginning and continuing French students found that, whilst there were significant differences in the proficiency of the two groups at Year 7 for some skills, these differences were no longer evident in Years 8 or 9 (Hill et al. 1997, Hill, 1998).

In Australia, whilst some researchers are enthusiastic about the cognitive and linguistic benefits of an early start to language learning (Clyne 1986; Clyne et al. 1995), others have expressed caution (Hill et al. 1997, Brown et al. 1999). One possible explanation for the conflicting findings is that expectations of higher achievement by primary-start learners are based on the assumption that there will be a smooth linear progression between primary and secondary LOTE study, but in practice there are relatively few secondary LOTE programs which stream learners according to whether or not they have studied the language in primary school and few schools provide an appropriate program for continuity of LOTE learning for those that have. In addition, the extent to which teachers are able to cater differently for learners in mixed classes is not known, as are the motivational effects of being in mixed Pathway classes.

### **1.3. Boys and girls as language learners**

Differences between girls and boys in language learning, both in achievement and in retention of language study, have long been a source of concern to educationalists. It has been found that girls are more likely than boys to continue with foreign language study at school, and it is commonly believed that they tend to do better than boys at learning languages. This belief is in line with the findings of studies investigating first language acquisition, i.e. that females generally outperform boys.

That female learners do better than males at learning foreign languages has been found to be the case in a number of studies. In a study of primary school French learning in England (Burstall, 1975) girls scores were found to be significantly higher on all tests. Ekstrand (1980) and Nyikos (1990) also report similar findings. Nykos focused specifically on vocabulary memorisation.

There are, however, some studies which report different findings. In a study of Hong Kong university students' on two tests of listening vocabulary, Boyle (1987) found that males had a higher average mean score on all ten tests administered, and on some tests the scores were significantly different. Scarcella and Zimmerman (1998) got the same result using a test of academic lexicon. Bacon (1992) reports no gender differences on listening comprehension scores.

Despite these mixed findings, the belief persists. Some researchers have attributed girls' superiority at learning to a more positive attitude to language learning (Burstall, 1975; Spolsky, 1989; Zammit, 1992). In Zammit's study of student attitudes, it was found that boys were less motivated than girls to learn a LOTE. They were more likely to say that they did not like learning languages, and reported more negative experiences in learning another language than girls did. Zammit also reports that a larger percentage of boys than girls found languages more difficult than maths. It has been claimed that these different attitudes result from differences in male and female culture (Malz and Borker 1982), in that learning a foreign language is a 'threat' to in-group male identity. Ellis (1994: 204) speculates that "the female 'culture' seems to lend itself more readily to dealing with the inherent threat imposed to identity by L2 learning".

Others argue that girls are better learners in general (Gass and Varonis, 1986; Bacon and Finemann, 1992). Oxford (1993) in a summary of research on language learning strategies, reports that females were generally found to use a broader range of strategies than males. Ellis (1994) suggests that the conflicting findings may

reflect both what learners know (where females are generally superior) and how they use the knowledge under given performance conditions (where males may prove superior). In other words boys may perform better on some tests not because they know more, but because they are better test takers.

#### 1.4. The effect of a home background in the language

Learners who have the opportunity to use the target language at home conversing with family members have obvious advantages in terms of the amount of target language input they receive compared with learners whose only target language exposure is in the classroom context. The issue of home use has been investigated in the context of 'heritage' language education in terms of language maintenance (Benyon and Toohey, 1991), and also bilingual education (Cummins, 1984), but to date little research has been done to examine the extent which home language use has an impact on 'foreign' language learning.

Elder (1996, 1997) compared the performance of background speakers and non background speakers on school foreign language examinations, and found a strong relationship between home exposure to the language and level of performance on the listening, and to lesser extent, the reading component of the test. However, Elder argues home use of the target language is not a sufficient condition for superior performance on school LOTE examinations especially when the language spoken at home is a dialect or a non-standard variety. She suggests that other factors such as level of home literacy, metalinguistic knowledge and experience of learning or studying languages other than the target language may also need to be taken into consideration in determining levels of success in formal language study.

## 2. Research questions

Three research questions were posed in relation to the issues discussed above:

- RQ1 Is student achievement comparable across different languages?
- RQ2 What is the effect of variables such as gender, L1 and home language use on achievement in school language programs?
- RQ3 What is the effect of the number of years of tuition on achievement in secondary school language programs?

### 3. Test development

In order to measure attainment and compare performance across the four languages chosen for this study – that is French, Indonesian, Italian, and Japanese – it was necessary to develop tests which were not only consistent with the curriculum focus of language programs in a diverse range of states and systems, but were also 'functionally equivalent' across the languages.

In order to do this, documents from all states and territories were reviewed in order to ensure that the views of proficiency espoused in the tests, and the range and level of skills, topics and tasks used in the test were consistent, as far as possible, with existing state and Commonwealth curricula and outcome statements.

Separate tests were prepared for each of Reading, Listening, Writing and Speaking by item writers with experience in test development as well as expertise in each of the four languages. Tasks were functionally-based and designed to cover a range of difficulty levels, topics and aspects of language knowledge. Test tasks (with English translations) were contributed by items writers from the respective languages for consideration. Tasks which were judged to be appropriate for all four languages were then translated into the other languages (from English). A complex process of fine-tuning across all four languages was undertaken in order to ensure that items remained functionally equivalent throughout the translation process. The tests were trialed at the end of 1998 and revised on the basis of statistical analysis *and* feedback from schools, interviewers, raters and outside experts.

The reading test consisted of a series of short texts accompanied by multiple choice questions. The listening test consisted of a series of short monologues and dialogues accompanied by multiple choice questions. The speaking assessment consisted of a 'live' interview containing two tasks. Task 1 was a set of formulaic questions asked by the interviewer to which learners were expected to respond orally. A score on a scale of 0-4 or 0-5 was awarded for each response. In Task 2 learners described a picture, and a score was awarded on a scale of 0-5. An overall assessment (on a scale of 0-3) was also made for Pronunciation and Fluency. The writing test consisted of a single task in which candidates were asked to write a short descriptive text. This was rated against three criteria, Task Fulfilment, Accuracy of Grammar and Spelling, and Resources of Expression, each rated on a scale of 0-4. Common criteria and descriptors were used for all languages for the writing and speaking tasks. What will be reported in the results section for *Task Fulfilment* are composite scores which

should be interpreted as general indicators of task-related and linguistic competence.

### 3.1. Quality of test data and reporting of scores

#### *Reading and Listening*

The tables below give the mean scores and standard deviations and measures of internal consistency for each language for Listening and Reading respectively.

	French	Indonesian	Italian	Japanese
Maximum score	30	30	30	30
Mean Score	19	18	18	18
SD	4.5	4.6	5	4
Internal consistency	0.78	0.78	0.81	0.7

Table 1: Testing quality – Listening

	French	Indonesian	Italian	Japanese
Maximum score	29	27	25	27
Mean Score	21	17	18	15
SD	4	4.6	3.5	5
Internal consistency	0.76	0.77	0.71	0.7

Table 2: Testing quality – Reading

#### *Speaking*

Two items within Task 1 were discarded before the analysis was undertaken as it was found that interviewers had frequently conflated the questions, resulting in a number of incomplete responses. The remaining eight items were analysed in order to evaluate whether it was appropriate to produce a single score by aggregating the scores for individual questions, that is whether performance on the questions was similar across all tasks (in measurement terms, whether unidimensionality exists in these data). A Rasch analysis of 'fit' was the method by which unidimensionality was investigated.

Findings across the four languages are reported in Table 3. Fit (which is, put simply, a measure of agreement between any given item and all other items, an aspect of test reliability) is expected to be 1.00, with measures of between 0.7 and 1.3 being acceptable. It can be seen that no measures are outside these values, leading us to conclude that it is



appropriate to add together the scores across all questions in Task One in order to produce a single score.

Question No.	<i>Japanese</i>	<i>French</i>	<i>Italian</i>	<i>Indonesian</i>
2	1.14	.96	.98	.72
3	1.07	1.12	1.15	1.03
4	.90	.94	.84	.88
5	1.06	1.01	1.05	1.15
6	.89	1.18	1.14	1.00
7	1.23	1.08	.92	.86
8	.83	.89	1.01	1.03
9	.95	.91	.87	1.00
Mean	1.01	1.01	1.00	.96
SD	.14	.11	.12	.13

Table 3: Unidimensionality of speaking data, Task One  
(Fit statistics)

Secondly, correlation between Task One and Task Two scores were investigated. As these were reasonably high across all languages (with correlation of between .73 and .96), a decision was made to reduce the two 'task' scores to a single score. This single score (reported as *Task Fulfilment*) can be interpreted as reflecting a general measure of linguistic (grammatical and lexical) skill.

It was not considered appropriate to conflate the scores for Task Fulfilment with the scores for Pronunciation and Fluency. A number of studies of test data have indicated that performance on each of these categories can be substantially different, i.e. that scores on linguistic competence do not necessarily correlate with scores for pronunciation or fluency, both of which are production skills rather than knowledge skills, but with fluency being also affected by, for example, candidate personality.

For this reason, three individual measures of proficiency are provided in subsequent analyses. The first measure, Task Fulfilment, is reported as a percentage score, and should be considered the main indicator of linguistic (grammatical and vocabulary) competence. Pronunciation and Fluency are reported on a scale of 1-3 and are less 'global' production-related aspects of speaking ability. There are nevertheless, some interesting questions relating to these criteria, such as whether, as has been suggested, students who have studied a language in primary school achieve substantially better pronunciation skills. These questions will be the focus of later analyses.

### Writing

The question of whether to add together the scores on the three writing criteria was considered next. Correlations between the three criteria were found to be significant, but as this may be simply a consequence of the high number of subjects involved, it was determined that the decision should be made on the basis of whether it made sense intuitively to add these criteria together. After consulting the raters, it was decided that it was appropriate to add the three together to produce a single score. This single score can be interpreted as providing a general measure of writing competence.

### 4. Data collection

As previously mentioned, the materials had been prepared to assess learners at the end of Year 8. However, as it was not possible to assess learners at the end of the year, the tasks were instead administered to students early in Year 9. The reading, writing and listening sections of the test were administered in class time by the regular LOTE teacher. Students were rostered to be withdrawn from class one at a time for the interviews, which were conducted by trained interviewers.

	French	Indonesian	Italian	Japanese	TOTAL
Reading	810	591	723	523	2647
Listening	810	598	738	599	2745
Writing	788	568	753	558	2667
Speaking	424	218	326	294	1262

Table 4: Numbers for each sub-test

Not all students completed all four test components. The numbers for each language and each skill are presented in Table 4. Because of the time required to conduct interviews with each student, the speaking component was administered to a selection of learners only.

### 5. Results

#### Is student achievement comparable across different languages? (RQ1)

Table 5 presents the scores in the four skills across the languages. There are three scores for *Speaking* and one each for *Reading*, *Writing* and *Listening*. Scores are reported as percentages for all skills except *Pronunciation* and *Fluency*, which were scored out of 3. Figure 1 presents the same information in graphic form.

Skill	French			Indonesian			Italian			Japanese		
	M	SD	N	M	SD	N	M	SD	N	M	SD	N
Speaking												
Task*	72%		408	75%		234	72%		311	63%		254
	13.2			14.8			17.0			18.8		
Pron.	1.89			2.5			2.24			1.91		
	.45			.66			.48			.38		
Fluency	2.1			2.28			2.07			1.91		
	.71			.8			.71			.78		
Writing	61%		786	69%		518	59%		735	52%		542
	19.8			20.6			20.0			21.9		
Reading	73%		793	64%		530	73%		722	56%		595
	13.9			17.2			13.8			19.2		
Listening	63%		791	59%		529	61%		711	59%		580
	14.9			15.5			17.0			14.2		

\*Task - Task fulfillment

Table 5: Comparison of mean scores across skills and languages

Fig. 1 Mean scores for Speaking, Writing, Reading and Listening

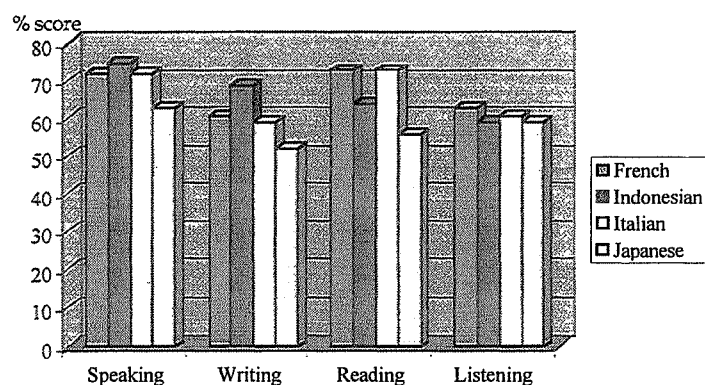


Table 6 shows the relative difficulty of the four languages for each skill in terms of mean scores. For example, for *Speaking (Task Fulfilment)*, the highest mean score was achieved in Indonesian ( $M = 75\%$ ), so this is ranked first, and the lowest in Japanese ( $M = 63\%$ ), so this is ranked last.

Learners of Indonesian achieved the highest scores on the productive skills, *Speaking* and *Writing*, and learners of Japanese generally the

lowest. Only in *Pronunciation* did a language other than Japanese achieve the lowest score, and that was French.

Scores on the reading tests were relatively disparate, ranging from 56 to 73%. French learners achieved the highest scores in *Reading* and *Listening*, followed by Italian learners, then learners of Indonesian. Learners of Japanese performed the least well. Scores on the *Listening* test, on the other hand, were relatively similar across the four languages, with French and Italian again ranked first and second.

	French	Indonesian	Italian	Japanese
Speaking				
Task fulfilment	3	1	2	4
Pronunciation	4	1	2	3
Fluency	2	1	3	4
Writing	2	1	3	4
Reading	1	3	2	4
Listening	1	3	2	3

Table 6: Ranking of performance by language and skill

	df	Mean Square	F	p	Effect size
Speaking					
Task fulfilment	3	7259.8	28.9	.000	.068
Pronunciation	3	22.8	91.2	.000	.19
Fluency	3	5.8	10.4	.000	.03
Writing	3	84290	227.9	.000	.21
Listening	3	2474	10.3	.000	.012
Reading	3	43363	171.6	.000	.16

Table 7: Comparison of the test scores across the languages

ANOVA statistics were performed to examine whether scores achieved in the different languages for each skill differed significantly. The results are summarised in Table 7. Differences were found to be highly significant on all four skills. Post-hoc tests revealed the Japanese scores to be significantly lower than those of other languages in *Speaking (Task Fulfilment)*, *Writing* and *Reading*.

What is the effect of variables such as gender, L1 and home language use on achievement in school language programs? (RQ2)

What is the effect of the number of years of tuition on achievement in secondary school language programs? (RQ3)

Research Questions 2 and 3 concerned the impact of variables such as gender, home language use and starting age on performance in the target LOTE. A questionnaire served to gather information about the demographics of the sample to assist us in our interpretation of the test data. The summary of the questionnaire survey is given in Tables 8 and 9.

Table 8 shows the distribution of the trial test sample across the four languages, according to gender, first language (L1) and language spoken at home (*Homeuse*). Across the four languages together, more than 70% of the population was girls. The proportion of boys was the largest in the Indonesian group (39.2%) and the lowest in French (17.6%)

Of the four language groups, the Italian cohort was most likely to have the target language as their first language and to use it at home. Both of these factors, *First Language* and *Home Use*, were considered to be likely influences on achievement. The impact of these variables on average achievement for the cohort as a whole, and the implications of this for the across-language comparison are discussed in the section *Language Background*.

Language	Total N	Gender		First language (L1)		Homeuse	
		Girls %	Boys %	Eng/ Other %	TL %	Eng/ Other %	TL%
French	847	82.4	17.6	99	1	99	1
Indonesian	568	60.7	39.3	99	1	99	1
Italian	791	72.4	27.6	95	5	94	6
Japanese	620	72.4	27.6	99	1	99	1

Table 8: Learner background variables

Table 9 shows the proportion of students according to whether they had or had not studied the same language at primary school. Learners of Italian were far more likely to have studied the same language in primary school than learners of the other three languages: over half of them (65%) had studied Italian in primary school, and of these 42% had studied it for 4 or more years at that level. In contrast, relatively few French (27%), Japanese (25%) and

Indonesian (20%) students had studied the same language prior to commencing secondary school. Overall, learners in the Italian cohort had also studied the target LOTE for an average of 4.8 years compared with 2 years for Indonesian, 2.1 for Japanese and 2.2 for French. In exploring the relationship with test score, the number of years of study, rather than starting age, was used in this study because the questionnaire data was more transparent in respect of this question. In general there was a relationship between the two, i.e. the starting age could be calculated by counting back from the present, but for some students this was not the case as learning had been discontinuous (e.g. a year in primary and then again in secondary, with a gap between). Such students were, however, few.

	Studied at Primary	%	Didn't study at Primary	%	Total
French	236	27.9	611	72.1	847
Indonesian	158	27.8	410	72.2	568
Italian	550	69.5	241	30.5	791
Japanese	187	30.2	432	69.8	619

Table 9: Primary language study

#### Differences in attainment amongst elements of variables

Statistical analyses were carried out to examine if variables such as gender, years of studying the language, home use (Italian group only) and learner's L1 (Italian group only) had any impact on the test performance. Non-parametric tests should ideally be used given that there were unequal numbers across the groups within each variable (language, gender, etc.) for some of the tests, and the variances lacked homogeneity. In reporting results here, both parametric (t-test or One-way ANOVA) and non-parametric (Mann-Whitney or Kruskal-Wallis) analyses are reported. The tables give an indication of sample size in each case.

#### *The effect of Gender*

Table 10 (see Appendix 1) presents the findings for the variable *Gender*, and Table 11 summarises the significance and direction of the differences. In French and Japanese, girls achieved higher scores on all skills, and differences were significant in *Speaking (Task Fulfilment)*, *Writing* and *Reading* in both of these languages. In Indonesian and Italian, on the other hand, boys generally did better than girls, except on Italian *Reading* and *Writing*. Some of the differences were significant.

	French	Indonesian	Italian	Japanese
Speaking				
Task fulfilment	girls	boys	boys	girls
Pronunciation			boys	
Fluency		boys	boys	
Writing	girls	boys		girls
Listening	girls	boys	boys	girls
Reading	girls			girls

n.b. Blank cells denote no gender differences in test score.

**Table 11: Gender differences in test scores**

*The influence of the LOTE as first or home language*

As mentioned earlier, there were more students with the target language as L1 and more students using the target language at home in the Italian group than in any other group. For this reason, it was considered necessary to investigate the performance of such 'background' speakers compared with non-background-speakers, and to evaluate the impact of any difference on mean scores for the whole Italian cohort, in order to ensure that comparisons across the languages were meaningful. (The number of background speakers studying French, Indonesian and Japanese was insufficient for statistical comparison of the two groups to be carried out.)

Descriptive statistics and the results of the statistical analyses for Italian are summarised in Table 12 (see Appendix 2). *Italian as L1* was found to have a highly significant effect on scores on all skills, as was *Homeuse of Italian*.

Although background learners performed significantly better, they are only a very small percentage of the overall cohort for that language (5-6%) and these variables are therefore likely to have only a limited impact on overall mean scores for Italian. For this reason it was considered reasonable to disregard their impact in the comparison across languages (as, indeed, it was for the other languages where it is even less of a factor in overall mean scores).

*The effect of years of study on test scores*

The effect of years of study on scores was investigated in relation to three groups: students with 0-2 years of study of the language (Group A), those with 3-5 years (Group B), and those with 6-8 years (Group C). The results of the analyses are presented in Table 13 (Appendix 3). Table 14 summarizes the significance and direction of the differences

across groups for each language, and the information is also presented graphically in Figures 2a – 2d.

	French	Indonesian	Italian	Japanese
Speaking				
Task fulfilment	C>A, B		A>B, C	
Pronunciation			C>A, B	
Fluency			C>A, B	
Writing	C>B>A	A>B, C		
Listening	C>A, B	B, C>A	C>B, A	
Reading	C>B>A	A>B, C	C, B>A	

n.b. Blank cells denote no differences in test scores in terms of Years of Study. A = 0-2 years, B = 3-5 years and C = 6-8 years, '>' denotes a significant difference (e.g., A>B Group A obtained a significantly higher score than Group B).

Table 14: Significant differences in test scores by years of study

The greatest gains from one group to another were found in French. As can be seen in Table 14, significant differences were found for French in all skills except *Pronunciation* and *Fluency*. In the Italian group, students who had studied Italian for 3 - 5 years achieved significantly higher scores in *Listening* and *Reading*.

There were very few students who had studied Indonesian for more than 5 years, so Groups B and C were combined. A significant difference was found between the two groups on *Listening*. Interestingly, learners who had studied Indonesian for less than 5 years received significantly higher scores for *Writing* and *Reading* than learners who studied for more than 5 years.

In Japanese, no significant effect was found for years of study on any skills. The failure to find significant differences is, of course, a consequence of the limited amount of progress learners make in the language compared to the diversity of proficiency levels within any one group.



Fig. 2a Differences in proficiency by Years of Study (French)

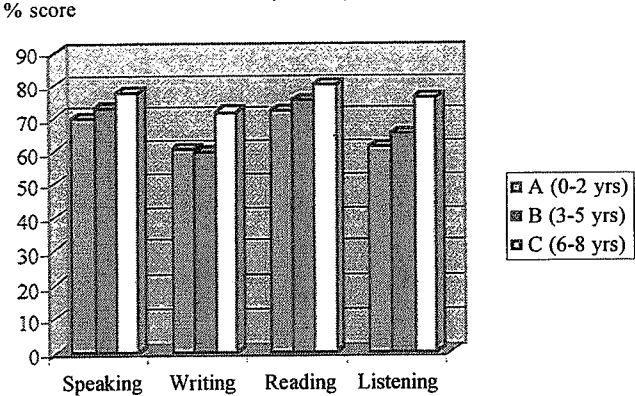
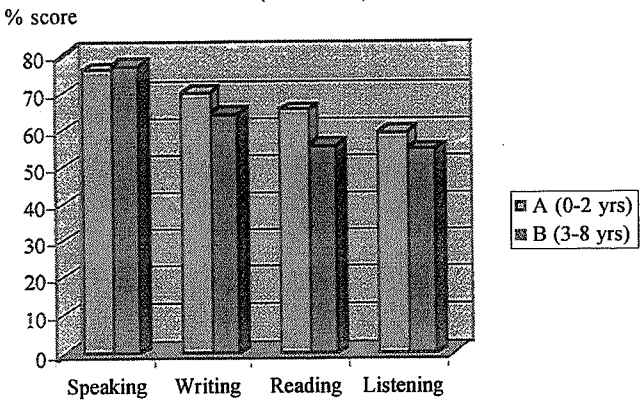
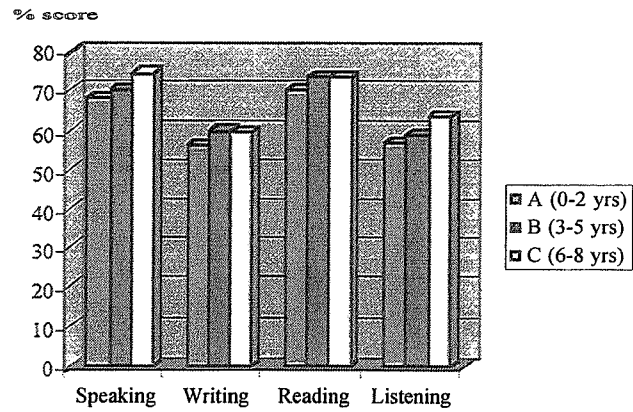


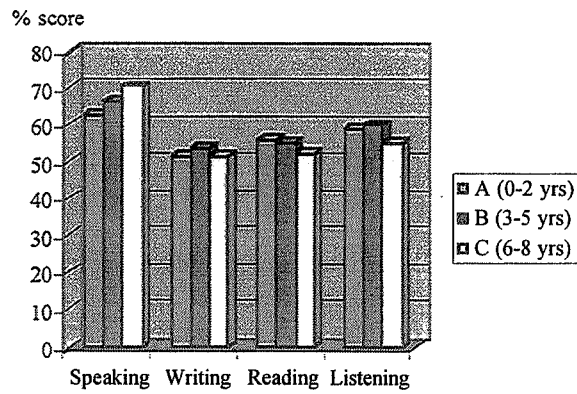
Fig. 2b Differences in proficiency by Years of Study (Indonesian)



**Fig.2c Differences in proficiency by Years of Study  
(Italian)**



**Fig. 2d Differences in proficiency by Years of Study  
(Japanese)**



Figures 2a-2d show that mean scores do not always rise steadily according to years of study as mentioned above. In several cases (*Writing* Indonesian, *Reading* in Indonesian, Italian and Japanese, and *Listening* in Indonesian and Japanese) mean scores actually decreased from Group B (3-5 years of study) to group C (6-8 years of study) for French, Italian and Japanese, and from Group A (0-2 years) to Group B/C (3-8 years) for Indonesian. What this means, in fact, is that on these skills and languages students who had studied the language right through primary school as well as secondary school (i.e. Group C) did not perform any better than students who had undertaken limited study (not all years) in primary school. This is reminiscent of the findings of Davies et al. (1997).

### Summary of analyses

The following are considered to be the main findings in relation to achievement across the current data set (a 'snapshot' of performance by Year 8 students in 1999):

- Levels of achievement are typically lower on all skills in Japanese than on the other three languages. Students of Indonesian achieved the highest scores in the productive skills (*Speaking* and *Writing*) but were outperformed by learners of French and Italian on the receptive skills (*Reading* and *Listening*).
- The number of years of study of a language has an effect on levels of achievement to a limited extent only. Not all skills increased significantly for all languages, and for Japanese, differences in levels of achievement across the three groups were not statistically significant in relation to any skill.
- Home language variables have a significant impact on levels of achievement across all skills in Italian. (It was not possible to investigate this issue for the other languages as the number of background learners was too small.)
- In general, girls outperformed boys in French and Japanese, but boys outperformed girls in Indonesian and Italian. Girls also seemed to generally perform better than boys on the literacy skills (reading and writing).

### 6. Discussion and conclusion

The analysis compared achievement across the four languages in the four skills. It also compared the achievement of boys and girls in each

language, of students with and without a home background in the language, and of students with different amounts of school exposure to the language.

That Japanese was found to be the 'hardest' language on all skills seems to support the language distance theory. For the literacy skills in particular, Reading and Writing, it is not surprising that learners of Japanese recorded the lowest achievement given that Japanese is a character-based script and learners are required to master three different scripts (Katakana, Kanji and Hiragana). It is also morphologically and syntactically more distant than the other languages, and has fewer cognates. These differences will affect all skills, of course.

That French and Italian, the two European languages, achieved the highest scores in the receptive skills would also seem to support the language distance theory. In the productive skills, however, Indonesian was found to be easier than the European languages. There are, of course, a number of possible explanations for this. It may be that Indonesian at such a basic level is, in fact, syntactically and morphologically simpler than the European languages, so that students perform 'better'. On the other hand, it may be that although raters were using the same scale for each language they were assessing student performance to different implicit standards (with, in this case, lower standards expected in Indonesian). Unfortunately, as all raters in this study were able to rate one language only, any across-language variation in standards was not perceivable or controllable. Again, for French pronunciation (which obtained the lowest scores), it may be that French pronunciation is intrinsically harder than that of the other languages, or it may be that French teachers expect higher standards or see good pronunciation as more critical than in the other languages (including Japanese).

A third possible explanation for learners of Indonesian achieving higher scores on the productive skills, and learners of French and Italian on the receptive skills (or indeed any other differences) is that teachers of different languages tend to emphasise different aspects of language skill in their teaching and this is reflected in student achievement. A detailed study of the teaching of different languages would shed light on whether, or the extent to which, different 'cultures' of teaching exist for different languages.

The findings of this study in relationship to language difficulty seem to support the use of language-specific standards, such as is the case in Victoria, rather than describing school-based foreign language learning standards in generic rather than language specific terms.

Schools, parents and students themselves need to be aware of potential differences in attainment as part of their decision-making processes (Which languages should we offer? Which language should we choose?) but also, perhaps more importantly, teachers should have realistic achievement targets to work to. The use of generic standards masks the differences across languages, and implies equivalence of difficulty.

The findings of this study for the variable *Gender* are very interesting. While they are consistent within languages, they are opposite in direction in different languages: girls consistently outperformed boys in French and Japanese, but boys outperformed girls in Indonesian and Italian. It is also interesting to note that differences were significant for the oral/aural skills (Speaking and Listening) in all languages, but when girls outperformed boys they also did so significantly on the receptive skills (Reading and Writing). It appears, then, that girls are more likely to perform well on literacy skills than boys.

It is difficult to infer any reason for the differences in performance across the four languages. It may be that certain languages are favoured more by boys, although whilst Indonesian has the highest proportion of boys studying it (39.3 % of the cohort), and French the lowest (17.6%), Italian and Japanese had exactly the same proportion (27.6%). It is also possible that the effect may be a consequence of the sampling. It is particularly difficult in studies such as this which rely on willingness by schools, teachers and students (and their parents) to participate, to ensure a balanced and representative spread of school and learner types and, although every attempt was made to ensure the samples for each language were similar, it may be that there was some bias in each of the languages towards particular high-achieving single-sex schools. Finally, it is worth noting that not all differences were significant, and even where they were they were not always substantial. Perhaps we should take note of comments made in relation to sex differences in L1 learning that beliefs about sex differences far outstrip reality (Marecek 1995: 105), and that "the similarity is much greater than the difference, and the difference itself may be an artefact of sampling" (Crawford, 1995:4-5).

When we investigated the relationship of home language variables with achievement for Italian, significant differences were found between the performance of learners with and without a home background in the language. However, as the proportion of background learners was relatively small (5-6%), there was likely to be little impact from this variable on achievement in the language across the whole cohort. Nevertheless, given the relationship between

language background and proficiency found here, one must question whether the language-learning needs of such students are adequately catered to within mainstream LOTE classrooms. The development of LOTE curriculum documents for background and non-background learners of languages such as Italian, Greek and Chinese within Victorian schools is a move to address this issue.

In relation to years of study, it was found that *most* learners of Italian had had considerably more years of study than learners of other three languages, with an average across the whole cohort of 4.8 years vis-à-vis 2 years for Indonesian, 2.1 for Japanese and 2.2 for French. On the basis of this, it would be expected that learners of Italian would demonstrate higher levels of proficiency overall. It is all the more surprising, then, to find that learners of Italian did not achieve the highest scores, although as noted earlier this is in line with a number of other studies. Given that a large proportion of the Italian cohort had studied the LOTE through primary school, the finding appears to bear out concerns expressed about the value of primary LOTE learning in relation to levels of achievement (although it should be noted that there are other potential benefits of primary LOTE learning which have not been addressed in this study). We offer one caution, also, in interpreting these findings: namely that such test-based data needs to be supplemented with an analysis of how teachers themselves cater for learners with different amounts of prior learning within mixed classes. We do not know to what extent teachers are equipped and prepared to be able to deal with mixed classes of primary and secondary start learners. Ultimately this is a problem for education departments to address in LOTE teacher development, perhaps.

This study has provided to opportunity to explore a number of contentious issues in school-based language learning achievement in Australia, drawing on a substantial database of student performance from schools around the nation. Whilst it adds to the existing literature on patterns of performance across languages and in relation to learner variables, the nature of survey studies such as this (in particular the difficulties of ensuring equivalent samples in relation to school and program type) leads us to suggest that additional studies would be advisable in order to substantiate the findings reported here. In addition, more qualitative studies of the learning context might shed some light on the reasons for such findings.

## 7. References

- Bacon, S. (1992) The relationship between gender, comprehension, processing strategies, and cognitive and affective response in foreign language listening. *Modern Language Journal* 76:160-178.

- Bacon, S. and Finnegan, M. (1992) Sex differences in self-reported beliefs about foreign language learning and authentic oral and written input. *Language Learning* 42: 471-495.
- Benin J. and K. Poohey (1991) Heritage language education in British Colombia: policy and programs. *Canadian Modern Language Review* 47: 4: 606-616.
- Blondie, CH, M. Candela, P. Edelenblos, R. Johnstone, A. Kubanek-German, Taeschner, T. (1998) *Foreign languages in primary and pre-school education: contexts and outcomes*. London: CILT.
- Boyle, J. (1987) Sex differences in listening vocabulary. *Language Learning* 37: 273-284.
- Brown, A., K. Hill, and N. Iwashita, (1999) *A longitudinal and comparative study: the attainment of language proficiency*. Final report submitted to Department of Education and Youth Affairs, Canberra.
- Burstall, C. (1975) Factors affecting foreign language learning: A consideration of some recent research findings. *Language Teaching and Linguistic Abstracts* 8: 5-25.
- Crawford, M. (1995) *Talking difference: On gender and language*. London: Sage.
- Cummins, J. (1984) *Bilinguals and special education: issues in assessment and pedagogy*. Clevedon, Avon: Multilingual Matters.
- Davies, A. and C. A. Elder, (1997) Language Distance as a Factor in the Acquisition of Literacy in English as a Second Language In P. Mackay, A. Davies, B. Devlin, J. Clayton, R. Oliver, and S. Zammit (eds), *The Bilingual Interface Project Report*. Canberra: DEETYA.
- Ekstrand, L. (1980) Sex differences in second language learning? Empirical studies and a discussion of related findings. *International Review of Applied Psychology* 29: 205-259.
- Elder, C.A. (1996) The effect of language background on 'foreign' language test language learning test performance: the case of Chinese, Italian, and Modern Greek. *Language Learning* 46, 2: 233-282.

- Elder, C.A. (1997) The background speaker of Italian, Chinese and Modern Greek: Implications for foreign language assessment. Unpublished PhD thesis, The University of Melbourne.
- Ellis, R. 1994: *The study of second language acquisition*. Oxford: Oxford University Press.
- Gao, M. (1996) Asian languages in Australia. *Babel* 31, 1, Apr-Jun, 14-17.
- Gass, S. and E. Varonis (1986) Sex differences in NNS/NNs interactions. In R. Day (Ed.) *Talking to learn: Conversation in second language acquisition* (pp.327-351). Rowley, MA: Newbury House.
- Genelot (1996) *L'enseignement des langues a l'école primaire: quels acquis pour quels effet au college? Elements d'évaluation: le cas d'anglais* (Les notes de l'Iredu). Dijon:Institut de Recherche sur l'Economie de l'Education.
- Genessee, F. (1994) (ed) *Educating Second Language Children: the whole child, the whole curriculum, the whole community*. Cambridge: CUP.
- Gibbons, J. (1994) Depth or breadth? Some issues in LOTE teaching. *Australian Review of Applied Linguistics* 17, 1: 1-22.
- Hyde, J. and M. Linn, (1990) Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin* 104: 53-69.
- Harley, B. (1986) *Age in Second Language Acquisition*. Multilingual Matters, Clevedon.
- Hill, K., A. Davies, J. Oldfield and N. Watson (1997) Questioning an Early Start: the transition from primary to secondary foreign language learning. *Melbourne Papers in Language Testing* 6, 2: 21-36.
- Hill, K. (1998) *A Comparison Of Beginning And Continuing Students Of French In Years 7, 8 & 9*. Report prepared for Presbyterian Ladies' College, Victoria and for the Association for Independent Schools Victoria. July 1998.
- Kenning, M. (1992) 'The Joint Languages Diversification Model: Aspects of a Case Study' *Language-Learning* 5, Mar, 2-5.



- Malz, D. and R. Borker, (1982) A cultural approach to male-female miscommunication. In J. Gumperz (ed) *Language and social identity* (pp.196-216). Cambridge: Cambridge University Press.
- Marecek, J. (1995) Psychology and feminism: can this relationship be saved? In Stanton, D, and A Stewart (Eds.) *Bilingual women: anthropological approaches to second language use* (pp.85-110). Oxford: Berg.
- Nyikos, M. (1990) Sex related differences in adult language learning: Socialization and memory factors. *Modern Language Journal* 74: 273-287.
- Odlin, T. (1989) *Language Transfer: Cross-linguistic Influence in Language Learning*. New York: Cambridge University Press.
- Oxford, R. (1993) Gender differences in styles and strategies for language learning: What do they mean? In J. Alatis (ed.) *Strategic interaction and language acquisition: Theory, practice and research* (pp.541-557) Washington, DC: Georgetown University Press.
- Ringbom, H. (1987) *The Role of the First language in Foreign Language Learning*. Clevedon: Multilingual Matters Ltd.
- Scarcella, R. and Zimmerman, C. (1998) Academic words and gender: ESL student performance in a test of academic lexicon. *Studies in Second Language Acquisition* 20: 27-49.
- Spolsky, B. (1989) *Conditions for second language learning*. Oxford: Oxford University Press.
- Willis, M. (1995) An illusion of difficulty or illusion: a reply to Andy Kirkpatrick, *Babel* 30, 2, Jul-Oct., 28-32.
- Zammit, S.A. (1992) The challenge: choosing to study a language other than English through high school. ASLLP Project No. 22.22/102. Melbourne: Australian Council for Educational Research.

## Appendix 1

	Girls			Boys			<i>t</i>	<i>p</i>	<i>W</i>	<i>p</i>
		Mean score	Mean rank		Mean score	Mean rank				
<b>French</b>										
Speaking-TF	332	73.4	206	64	67.5	159.5	2.6	**	8125	**
Speaking-Pro	336	1.9	204.5	66	1.8	186.2	1.3	ns	10077	ns
Speaking-Flu	332	2.1	202	66	2.0	186.2	1.1	ns	10078	ns
Writing	630	7.5	397	130	6.67	318.5	3.1	**	32893	***
Listening	645	21.9	410	129	18.7	274.9	6.2	***	27084	***
Reading	638	19.4	406	131	16.6	279.9	5.69	***	28024	***
<b>Indonesian</b>										
Speaking-TF	128	74.7	103.5	95	78.5	123.5	-2.04	*	4988	*
Speaking-Pro	128	2.44	107.1	95	2.57	118.6	-1.49	ns	5453	ns
Speaking-Flu	128	2.22	104.1	95	2.45	122.58	-2.26	*	5074.5	*
Writing	304	8.19	246.5	204	8.52	266.4	-1.48	ns	28589	ns
Listening	321	17.3	245.4	199	18.4	284.8	-2.7	**	27104	**
Reading	314	17.49	264.4	208	17.06	257.1	1.03	ns	31745	ns
<b>Italian</b>										
Speaking-TF	252	71.3	147.6	50	75.9	171.3	-1.76	ns	5309	ns
Speaking-Pro	252	2.2	146.0	50	2.44	179.1	-2.75	**	4918	***
Speaking-Flu	250	2.01	144	49	2.36	180.7	-3.22	***	4622	**
Writing	534	7.2	371	188	6.77	334	2.07	*	45144	*
Listening	506	17.7	333	193	19.4	394	-3.78	***	40209	***
Reading	518	18.17	356.4	193	18.1	365	.271	ns	49802	ns
<b>Japanese</b>										
Speaking-TF	179	65.6	127.8	63	59.7	103.6	2.24	*	4513	*
Speaking-Pro	176	1.79	122.5	63	1.85	112.9	1.52	ns	5098	ns
Speaking-Flu	174	2.06	128.2	62	1.61	91.38	4.03	ns	5712	*
Writing	401	6.7	294.6	137	5.1	196.2	6.7	***	17423	***
Listening	408	18.2	301.2	158	16.7	237	4.1	***	25018	***
Reading	423	15.9	319.9	162	12.8	222.8	7.8	***	22892	***

TF – task fulfillment, Pro – Pronunciation, FI – Fluency

\*  $p < .05$  \*\*  $p < .01$  \*\*\*  $p < .001$ 

Table 10: Impact of the gender variable on the test score

## Appendix 2

	Eng/Other	Mean score	Mean rank	TL	Mean score	Mean rank	<i>t</i>	<i>p</i>	<i>W</i>	<i>p</i>
<b>L1</b>										
Speaking-TF	268	70.05	141.2	29	85.32	221.2	-4.54	***	1791	***
Speaking-Pro	268	2.19	141.5	29	2.72	218.7	-5.92	***	1864	***
Speaking-Flu	265	2.01	141.6	29	2.55	201.6	-3.88	***	2273	***
Writing	674	6.9	348.4	36	8.7	487.8	-4.09	***	7369	***
Listening	653	17.8	332.7	35	24.5	564.7	-7.8	***	3719	***
Reading	667	18.0	342.4	37	21.05	534.9	-5.3	***	5590	***
<b>Home use</b>										
Speaking-TF	270	70.6	141.4	26	85.8	221.8	-4.45	***	1604.5	***
Speaking-Pro	270	2.2	141.1	26	2.81	225.9	-6.6	***	1498.5	***
Speaking-Flu	268	2.01	141.5	25	2.64	205.5	-4.24	***	1886.5	***
Writing	659	6.98	343.6	45	8.6	483.5	-4.46	***	8932	***
Listening	637	17.85	328.5	45	23.4	525.4	-7.2	***	6059	***
Reading	655	17.9	338.5	44	20.8	520.8	-7.8	***	6893	***

TF – task fulfillment, Pro – Pronunciation, Fl – Fluency

\*  $p < .05$  \*\*  $p < .01$  \*\*\*  $p < .001$ 

Table 12: Impact of L1 on the test score in Italian group

## Appendix 3

	0-2 yrs			3-5 yrs			6-8 yrs			<i>f</i>	<i>p</i>	<i>X or U*</i>	<i>p</i>
	<i>n</i>	Mean score	Mean rank	<i>n</i>	Mean score	Mean rank	<i>n</i>	Mean score	Mean rank				
<b>French</b>													
Speaking-TF	258	70.8	173.2	79	73.2	194.2	28	77.8	241.8	4.0	*	11.8	**
Speaking-Pro	259	1.8	181.3	82	1.9	186.7	28	2.0	214.5	2.20	ns	4.9	ns
Speaking-Flu	256	2.1	182.4	81	2.1	183.4	28	2.1	187.4	.028	ns	.064	ns
Writing	580	7.3	383.2	154	7.2	372.5	41	8.7	514.7	6.7	***	14.5	***
Listening	586	21.0	375.7	152	21.9	414	41	23.3	506.1	5.15	**	15.1	***
Reading	586	18.5	366.5	150	19.6	418.7	41	23.0	602.1	22.1	***	45.75	***
<b>Indonesian</b>													
Speaking-TF	192	76.3	110.6	29	77.4	113.7				-.40	ns	2075	ns
Speaking-Pro	192	2.52	112.8	29	2.41	102.2				.808	ns	2528	ns
Speaking-Flu	192	2.30	110.4	29	2.37	115.2				.790	ns	2661	ns
Writing	408	8.46	261.0	96	7.74	216.3				2.83	**	16107	**
Listening	428	16.6	269.0	94	17.8	227.2				-2.05	*	16888	*
Reading	420	17.7	263.4	88	15.8	211.9				3.63	***	14728	***
<b>Italian</b>													
Speaking-TF	96	18.81	137.3	62	17.75	144.8	142	15.2	161.9	4.02	*	4.9	ns
Speaking-Pro	95	2.17	140	62	2.18	140.6	143	2.32	161.8	3.6	*	7.83	*
Speaking-Flu	94	2.01	142.5	62	1.9	130.4	141	2.18	161.6	3.83	*	7.7	*
Writing	245	6.8	343.9	168	7.2	373.8	309	7.2	388.8	2.2	ns	2.77	ns
Listening	237	17.26	308.3	154	17.8	334.1	305	19.2	387.0	9.67	**	21.57	***
Reading	241	17.6	326.9	158	18.5	375.5	313	18.4	369.7	4.8	**	7.71	*
<b>Japanese</b>													
Speaking-TF	183	63.1	117.6	49	66.6	128.1	9	70.6	151.1	1.3	ns	2.623	ns
Speaking-Pro	181	1.91	118.5	48	1.94	121.7	9	2.0	128	.282	ns	.654	ns
Speaking-Flu	179	1.91	121.6	47	1.74	101.7	9	2.11	131.4	2.15	ns	4.05	ns
Writing	410	6.24	265.4	98	6.5	278.8	28	6.32	277.8	.387	ns	.704	ns
Listening	426	17.7	282.4	110	18.1	297.9	30	16.7	245.7	.450	ns	2.484	ns
Reading	444	15.1	294.0	111	15.1	296.4	30	14	265.3	.062	ns	.873	ns

\*For Indonesian Group, there were only two groups, and therefore Mann-Whitney Test (*U*) was performed.

*p* < .05 \*\* *p* < .01 \*\*\* *p* < .001

Table 13: Impact of Years of Study on the test score