Chapter 5
The Analysis of the Learners' Test Results

This chapter presents the analysis of the data collected on the test results of the Grade 3 learners in 19 villages. The analysis examines the relationship between language choices in the class and Grade 3 learners' achievements. The statistical analysis is done with great help by Dr. Robert Wyn Owen.

5.1 Overview of the data

This section provides an introduction to the data on the learners' test results. This study utilizes the data on Grade 3 learners' test results from 19 classes which represent about one third of the total number of 54 schools in the target area. The monthly, chapter-end test scores from November 2009 are available for the Burmese; Mathematics; Science and English subjects. Table 2 gives a summary of the data.

Each village and school is given a number from 01 to 19 and the particular class is labeled according to the mother tongue of the Grade 3 teacher in Nov 2009. So, for example, class B01 denotes the Grade 3 class in school 01 whose teacher speaks Burmese as her mother-tongue whereas class S04 denotes the Grade 3 class in school 04 whose the mother tongue is Sgaw. The three-letter codes in the 'Teacher Language' column indicate the pattern of languages of instruction used by the teachers who taught the present Grade 3 class while they were in Grades 1, 2 and 3. So, for example, SSB denotes that in Grades 1 and 2 the teachers were both Sgaw mother tongue speakers whereas the Grade 3 teacher was a Burmese mother tongue speaker. The final four columns give typical marks (as measured by the median for each class) for the four subjects on which the learners were tested, namely Burmese, Mathematics, Science, and English.

Table 2 Summary of the data

Class Label	Teacher Language	Class size	No. of Males	No. of Females	Burmese (Median)	Mathematics (Median)	Science (Median)	English (Median)
B01	SSB	23	14	9	50	64	64	66
B02	SSB	24	13	11	62.5	77	65	75
B03	SSB	25	21	4	<40	<40	45	42
S04	SSS	26	9	17	90	92	94	96
B05	SSB	9	6	3	68	80	74	80
S06	SSS	31	19	12	46	44	44	64
B07	BBB	38	22	16	42	58	52	72
S08	SSS	18	11	7	84	66	84	70
B09	SSB	10	7	3	46	44	70	50
S10	SSS	24	10	14	87.5	91	92	60
S11	SSS	12	4	8	76	73	82	92
S12	SSS	23	11	12	68	49	59	64
B13	SSB	18	9	9	66	52	50	79
S14	SSS	38	24	14	59	52	60	<40
S15	SSS	4	3	1	92	50	76	43
S16	SSS	24	13	11	80	74	80	72
S17	SSS	24	11	13	82	92	70	76
S18	SSS	10	3	7	88	78	92	74
B19	SSB	11	6	5	84	98	78	96
	Total:	392	216	176				

The data in Table 2 shows that there are only three different patterns of teacher language: SSS; SSB and BBB. There are 11 classes with the SSS pattern; 7 with SSB; and 1 with BBB. By totaling the appropriate columns it can be seen that there are 216 males, 176 females giving a total of 392 learners from 19 schools. The males make up 55% of the total sample and females 45%. Class B07 is the only one with the BBB teacher language pattern. We note that class B03 has the lowest marks of any of the schools, with over half the class failing the original tests in Burmese and Mathematics. One other aspect of B03 that stands out is the ratio of males to females; which is over 5:1. If there are differences in test performance for males and females, class B03 would be particularly affected. There is further discussion of the problems caused by a large proportion of failures in Section 5.2.3.

5.2 Test System

This section presents the test system and is divided into three sub-sections: test procedure; test papers; and test scores.

5.2.1 Test procedure

Learners are tested at the end of each month on the material studied that month. This is called a chapter-end test. According to the education system, in primary school, every student is supposed to move on to the next grade each year so they provide supplementary tests if the child fails the test the first time. Although officially Grades 1 and 2 learners can move on to next grade without passing the test, the teachers arrange the tests to give the learners motivation for wanting to try hard in the class.

Those learners who fail the original test (i.e. get a mark of less than 40%) are tested again. If they can't pass even the supplementary test the teachers have to check the learners records and discuss with their parents to make a decision on whether the student should move on to the next grade or not.

5.2.2 Test papers

The English translations of the test papers along with the marking schemes for each subject are given below.

Name of School
Assessing Achievement Exam
November 2009

Grade 3 **Burmese** **Time (30 minutes)**

1. Put the punctuation where it is necessary in the text given. **(5 marks)**

   For this question, a paragraph of three to four sentences is provided without any punctuation such as marks between phrases or marks at the end of the sentences. (Burmese has only two punctuation marks, roughly equivalent to a comma and a period in English.) The learners are to put those marks in the paragraph.

2. Fill in the blanks. **(4 marks)**

   a. The bird received by the monks is _____.
   b. Bees are living by _____.
   c. When you play "htotesito" the leader is called _____.
   d. In the battle for freedom, _____ participated.

3. Make a sentence using each word given below. **(8 marks)**

   a. elder b. rule c. happily d. martyr

4. Answer the questions. **(8 marks)**

   a. What is "atakuchin"?
   b. Describe the types of bees in a bee hive.
   c. What benefit can you get by playing "htotesito"?
   d. What is the date that Myanmar became independent?

Name of School
Assessing Achievement Exam
November 2009

Grade 3 **English** **Time (30 minutes)**

1. Choose the correct word given in the brackets. **(5marks)**

   a. The boys are playing (in, under) the tree.

   b. The cow is (in, under) the tree.

   c. Su Su is (behind, under) the door. She is hiding.

   d. The blackboard is (behind, in front of) the class.

   e. The picture is (in, on) the wall.

2. Look at the pictures and write the words. **(5marks)**

   a.  _____

   b.  _____

   c.  _____

   d.  _____

   e.  _____

3. Fill in the blanks with 'is' or 'are'. **(5 marks)**

   a. Where _____ the apple?

   b. Where _____ the eggs?

   c. The stars _____ in the sky?

   d. The orange _____ beside the apple.

   e. There _____ many boats in the river.

4. Look at the pictures and answer the questions. **(10 marks)**

   a. Where is the flower? It is _____.

   b. Where is the ball? It is _____.

   c. Where is the spoon? It is _____.

   d. Where are the two apples? They are _____.

   e. Where are the three glasses? They are _____.

Name of School
Assessing Achievement Exam
November 2009

Grade 3 **Mathematics** **Time (30 minutes)**

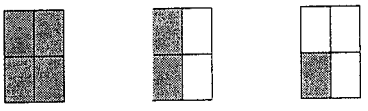
1. Choose the correct answers. **(5 marks)**

   a. Three umbrellas at the same price cost 96 so one umbrella will cost

      i. 32 ii. 33 iii. 34

   b. The fraction, 1/4 is

      i. ii. iii.

   c. 3/4 - 1/4 =

      i. 1/4 ii. 2/4 iii. 3/4

   d. 3 kyats 75 pyas + 6 kyats 25 pyas is

      i. 10 kyats ii. 11 kyats 10 pyas iii. 9 kyats 90 pyas

   e. 18 kyats 75 pyas – 9 kyats 25 pyas is

      i. 9 kyats 25 pyas ii. 9 kyats 50 pyas iii. 9 kyats 75 pyas

2. Fill in the blanks. **(5 marks)**

   a. You can subtract 2 from 8 _____ times.

   b. One out of 2, one half, and one part of a two sector circle is written in fraction as: _____.

   c. The numerator in 3/5 is _____.

   d. The denominator in 3/4 is _____.

   e. 3 kyats 25 pyas is equal to _____ pyas.

3. Calculate the sum. (a, b, c, d, e) **(15 marks)**

   a. Divide 936 by 3.

   b. If 12 oranges are kept in a bag, how many bags are needed for 927 and how many oranges are left.

   c. Calculate 1/4 +1/4 by drawing the diagrams.

   d. Calculate 3/4- 1/2 by drawing the diagrams.

   e. If you bought a hoe for 160 kyats 50 pyas, a shovel for 112 kyats, a knife for cutting grass for 66 kyats 75 pyas and a sickle for 80 kyats 90 pyas, how much did you pay in total.

Name of School
Assessing Achievement Exam
November 2009

Grade 3 **Science** **Time (30 minutes)**

1. Choose the correct answer. **(5 marks)**

a. When we hear the marching sound, we feel _____.

i. afraid ii. excited iii. weak

b. Rubbing two sticks of bamboo together can cause _____.

i. heat ii. light iii. magnet

c. In the day time, natural light comes from (the) _____.

i. candle ii. electricity iii. sun

d. At our age, we must find _____.

i. money ii. education iii. religious teaching

e. If you get pocket money, you should buy a snack that is _____ and makes you full.

i. nutritious ii. with toy iii. wanted

2. Answer the questions (short answer). **(8 marks)**

a. Describe a factor that can cause forest fires.

b. State two kinds of pleasant sounds that you like.

c. Write down two good deeds that you could do.

d. What is unclean food?

3. Answer the question (long answer). **(12 marks)**

a. Describe two kinds of sounds that make you feel pleasant and two kinds of sounds that make you feel unpleasant.

b. If you got pocket money what kind of snack must you buy and eat. Give two examples.

c. Is it good manners to hurt other people? Describe two kinds of actions that you should avoid at your age.

Note that the math questions refer to kyat and pya (100 pya = 1 kyat). Although pya coins were previously used in Myanmar, due to inflation they are no longer used and currently the smallest denomination currency is a 5 kyat note.

From the structure of the question papers, total scores might be expected to have a multi-peaked distribution. For example, if learners understand how to answer a particular question, they can then get marks from several parts of that question. However, if a student does not understand what is required in a particular question, they will not be able to answer any of the parts of that question and so all of the marks for that question will be beyond their reach.

5.2.3 Test scores

The test score data are recorded as percentages, but the tests were originally out of 25 and marks were multiplied by 4 to obtain a percentage. Half marks are apparently awarded since percentages of e.g. 50 and 42 are not divisible by 4. The fact that the original marks total was 25 will be of practical importance when interpreting differences between different groups. For example, a difference of 8% between two scores would be equivalent to two questions worth one mark each in the original test.

If a student passes the re-test, they are awarded a 40% in the mark list, no matter how much greater than 40% their score in the re-test. In the original test, if somebody scores 40%, their mark is recorded as 42% to show that they passed the original. It is thus always possible to identify which marks are from the retest. Marks of less than 40% in the recorded mark list could be either from the retest or from the original test if the student did not take the retest. In the analysis that follows it will not matter which test they are from.

The results show that many children, particularly boys, fail the original test. This is shown by the dramatic peak at 40% in the dotplot of all male scores in Burmese in Figure 5.

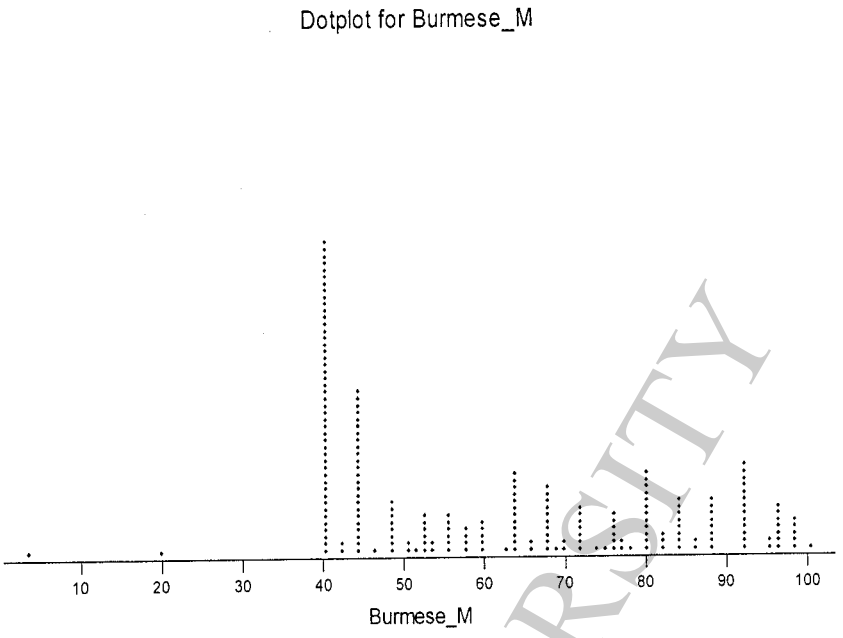


Figure 5 Dotplot for all male scores in Burmese

Schools do not keep records of the original scores for those learners who need to take the re-test. The data provided is therefore a kind of mixture distribution. The mean of such a mixture distribution is not of interest because it combines results from two different tests. Where a measure of typical score for a class is required, the median will be used, because the median of the mixture distribution is the same the median of the original distribution as long as the number of failures in the particular subgroup for which the median is calculated is less than half the group. For example, the number of males who fail Burmese is less than half the total males, so the overall median score for male learners in Burmese is the same as the original. However, for school B03 in Burmese, there are 17 failures out of 21 boys in the class. So the median score of the male learners in Burmese for B03 is not the same as the median of the scores of the original test. The median of the original B03 scores was probably much lower. The following sections explore factors that might explain patterns in the data.

5.3 Description of Factors

There are two types of factors: those that could not be measured and those that were measured. These are described in the following subsections.

5.3.1 Unmeasured factors

It was not possible to obtain appropriate data to directly assess the effects of the following factors, both of which might cause differences between classes: teacher proficiency; environmental factors.

5.3.1.1 Teacher proficiency

Although no actual numbers were collected, the author's experience as a member of the Sgaw community which is similar to the Sgaw people from the target community is that there are differences between the Sgaw teachers and the Burmese teachers. The primary difference is that Sgaw teachers have Sgaw as their L1; Burmese as their L2 and English as their L3. The Burmese teachers have Burmese as their L1 and English as their L2. Sgaw teachers are often from local villages so have limited exposure to Burmese (outside of their own education) and even more limited exposure to English. The Burmese teachers on the other hand speak Burmese as their L1 and, since they are from the larger towns, have had more exposure to English. As far as teaching is concerned, the Burmese teachers teach their L1 and L2 as subjects whereas the Sgaw teachers teach their L2 and L3. Another difference between the Burmese and Sgaw teachers is their level of education. The Burmese teachers have often studied at higher levels, usually having a bachelor's degree whereas the Sgaw teachers often do not have a bachelor's degree. Most of the Sgaw teachers in the target area are the community village teachers who are the ones the village appointed to fill a real need for teachers.

5.3.1.2 Environmental factors

The descriptions in Chapter 4 give details about the general situation in most of the villages. All of the villages are relatively poor and parents often have little education themselves so are unable to provide much help to their children with school work. Extra tuition where learners pay for help with schoolwork outside of school is only available in a few of the villages. Some of the villages have more contact with Burmese-speaking villages than others. Such exposure to Burmese would be expected to benefit children in their study of Burmese and their understanding of other aspects of the curriculum, although the exposure might not be sufficient to make any real difference in practice.

5.3.2 Measured factors

When analyzing the data it is necessary to consider all factors that might contribute to the pattern of tests scores observed. The focus of this study is the effect on the test scores of the language of the teacher. In order to claim that differences in test scores are due to differences in the language of the teachers of different classes, other factors must be accounted for. The data collected allows the following factors to be assessed directly: Class effect; Class size; Grade 3 teacher language. Grade 1 and Grade 2 teacher language were recorded for each class, but only one school, B07, had the BB pattern, all other schools have the SS pattern. There is therefore not enough data to estimate the effect of the BB pattern. In subsections 5.3.2.1 to 5.3.2.5 these measured factors will be explored one at a time, to check for evidence of correlations between each individual factor and test scores. In order to be concise, the explorations are only presented for one subject, namely Burmese. In Section 5.4, full analyses of test scores for all four subjects are presented. The explorations presented in the following subsections are not conclusive, rather they suggest potential relationships in the data which will be more rigorously examined in Section 5.4.

5.3.2.1 Class

Figure 6 shows boxplots of Burmese test scores for each class. The width of the boxes are proportional to the number of learners in the class, so the narrowest box is S15 which has 4 learners and the widest box is B07 which has 38 learners. The bold dots represent the median values for each class.

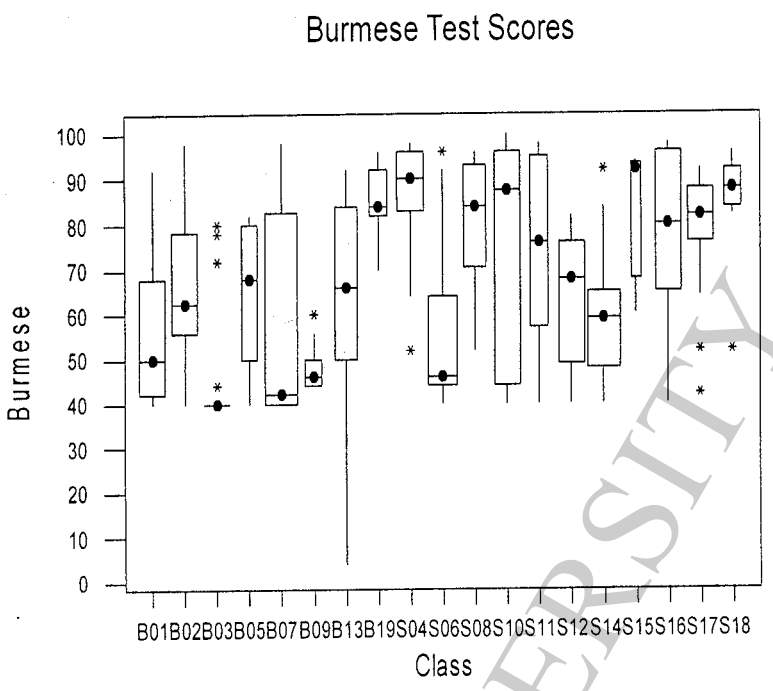


Figure 6 Boxplots of scores in classes

Figure 6 shows that median scores in classes with a Burmese teacher – grouped towards the left of the plot – are on the whole lower than the median score for classes with a Sgaw teacher. The median values vary greatly – from 40 for class B03 to 92 for class S15. The lengths of the boxes show that the distribution of marks within a class also varies considerably from class to class. The asterisks indicate scores that are extreme for a particular class. In other words, given the spread of the central 50% of the scores for a particular class (as depicted by the box) these scores are unexpectedly far from the nearest quartile (represented by the upper and lower edges of the box) and are considered extreme. There appears to be some correlation of class size with test score – for example, the high median score of the smallest class S15. (The narrowest box indicates it has the smallest class size.) However there are exceptions – for example the low median value of class B09. These observations suggestion that the particular class a student is in has an effect on test score, apart from the effects of Grade 3 teacher language and class size.

Since the true median value of the original test scores for class B03 is almost certainly much lower than the median (40%) of the mixture of both original test and re-test, this class is omitted from the following analysis. This is to prevent one

extreme class from having too much influence on general conclusions that are drawn from the data.

5.3.2.2 Grades 1 and 2 teacher language

It is to be expected that the effect on student's learning of not having a language in common with their teacher would be most acute in Grade 1 followed by Grade 2. It is therefore important to look at the effect on test scores of teacher language in Grades 1 and 2. As it turns out, the only class in the entire sample to not have Sgaw teachers in Grades 1 and 2 is class B07, which had Burmese teachers in both Grade 1 and Grade 2. This is probably due to the fact that there is a Burmese village in very close proximity to the village in which the school is located. There are even some Burmese-speaking children in the class.

Figure 7 shows median scores in Burmese for 18 classes differentiated by teacher language in Grades 1 and 2.

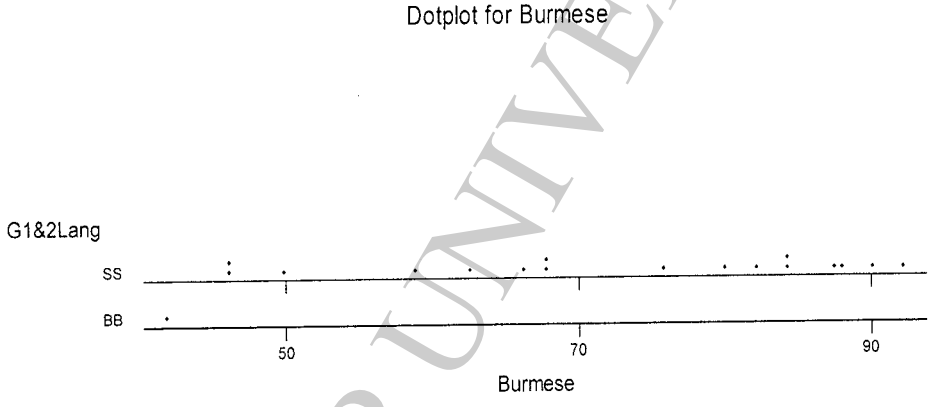


Figure 7 Dotplot for Burmese median (by teacher language in Grades 1 and 2)

As can be seen from Figure 7, the one class that had Burmese teachers in Grades 1 and 2 had the lowest median score in Burmese of the 18 classes. It should be noted that class B07 had the equal largest class size, namely 38 learners. While it is not possible to draw firm conclusions from just a single class, these results are consistent

with what one would expect: if learners do not have a language in common with their teacher in Grades 1 and 2 it has a severe negative impact on student learning.

In the analyses that follow, class B07 is omitted to ensure that any teacher language effect is due to the Grade 3 teacher language.

5.3.2.3 Class size

The relationship between class size and test performance is visualized in Figure 8 of class median scores in Burmese versus class size. Figure 6 also indicates the language of the teacher for each class.

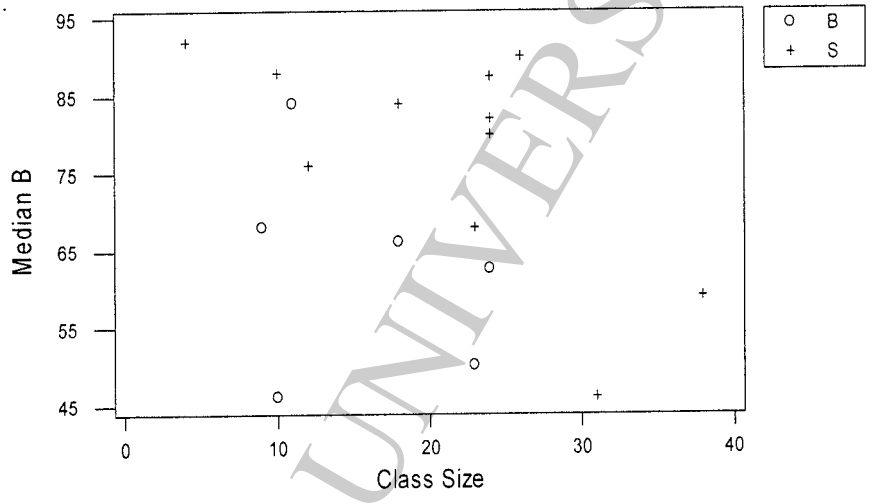


Figure 8 Burmese median scores (by class size)

As can be seen from the plot, higher median scores tend to be associated with lower class sizes and lower median scores tend to be associated with higher class sizes. This pattern is true for both the scores from classes with a Sgaw teacher (symbolized by a cross) and classes with a Burmese teacher (symbolized by a circle). This is what one would expect as for a larger class the teacher's attention has to be divided between a greater numbers of learners. Another point to note from the plot is that for a given class size, the scores from classes with a Sgaw teacher are generally higher than the scores from classes with a Burmese teacher.

5.3.2.4 Gender

There are differences in social roles for girls and boys in the village communities where this study was based. Outside of school hours boys might be asked to help parents in the fields whereas girls stay at home possibly looking after younger siblings but generally having more opportunity to study than the boys. Gender is therefore an obvious factor to examine in performance.

The differences between male and female scores in Burmese are plotted in Figure 9 for each class. This ensures the comparison between typical male and female scores is done while other factors such as class size and teacher language are controlled.

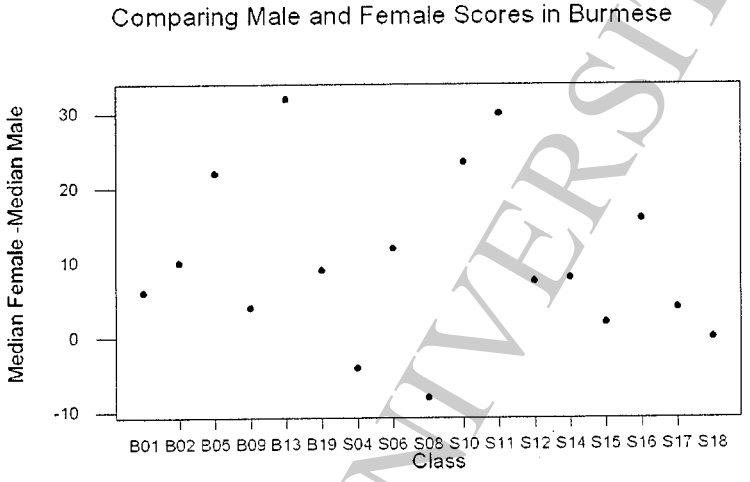


Figure 9 Difference in median Burmese test scores (Female minus Male)

As can be seen in the plot, the differences between the median score of male and female are shown on the vertical axis. Only two of the differences (for classes S04 and S08) are less than zero, meaning that in the remaining 15 classes the typical score for female learners is at least as great as that for the male learners. This suggests that gender is an important factor in test performance.

5.3.2.5 Grade 3 teacher language

There have been indications of the effect of Grade 3 teacher language in the plots in previous sections, for example in Figure 8 the median score for a class with a Sgaw teacher was generally higher than for a class of the same size with Burmese teacher.

Figure 10 compares the cumulative distributions of test scores for learners with Burmese and Sgaw-speaking teachers. For a particular score on the Burmese test (shown on the horizontal axis) the corresponding value on the vertical axis shows what proportion of the students achieved scores less than that value. For example, the dotted line representing the students with Sgaw teachers shows that approximately 25% of those students (vertical axis) achieved scores below 50% (horizontal axis). Both graphs end at the same point because all test scores (on the horizontal axis) are less than or equal to 100% and on the vertical axis is plotted on a percentage scale.

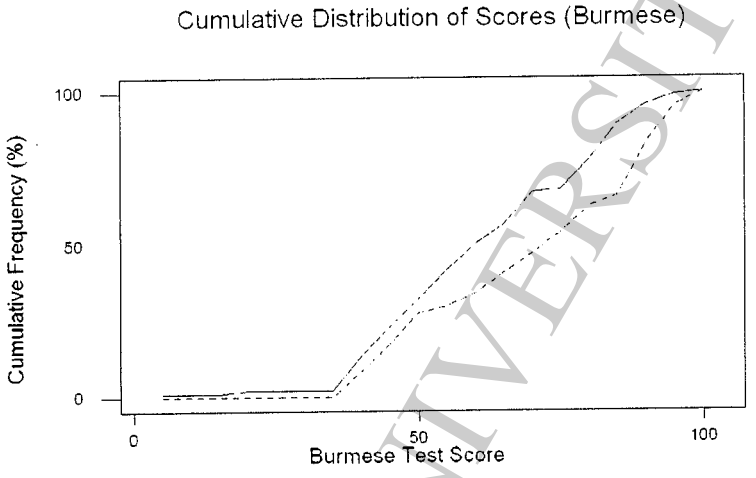


Figure 10 Cumulative percentage plot for Burmese (by teacher language in Grade 3)

As can be seen from Figure 10, the fact that the solid line representing the scores of learners with a Burmese-speaking teacher is always above the dotted line representing learners with a Sgaw-speaking teacher shows that the proportion of learners with a Burmese-speaking teacher getting marks below a certain value is always higher than for learners with a Sgaw-speaking teacher. In other words, learners in Sgaw classes generally get higher marks.

5.3.3 Summary of factors

To summarize, when examined one at a time, there is evidence that gender, Grade 3 teacher language, class, and class size are all associated with differences in test scores in the Burmese subject. The analysis that follows examines the ability of

combinations of factors to predict the test scores. As will be seen, there are different patterns between the different subjects, so each will be presented separately.

5.4 Analysis of Test Scores

This section presents the results of fitting statistical models to the data. The models help to show the relative importance of each factor in predicting test scores. To rule out the effect of Grades 1 and 2 teacher language, class B07 was not included in the following analyses. Class B03 was also omitted because of its extremely high proportion of failures. To overcome the problems caused by lack of normality in the test scores, the test scores were first ranked and the models were fitted to the ranked data.

5.4.1 Analysis of Burmese test scores

The best model for ranked Burmese scores was a mixed effects model including a random effect for class plus fixed effects for Gender, Class Size, and Grade 3 teacher Language. The relevant output from the R computer package is displayed in Figure 11.

```
Linear mixed-effects model fit by REML

Random effects:
 Formula: ~1 | class
        (Intercept) Residual
StdDev:    41.80242 80.04392

Fixed effects: brank ~ gender + n + TL
               Value Std.Error  DF   t-value p-value
(Intercept) 175.34661 30.250000 311  5.796582  0.0000
gender       30.90018  9.019417 311  3.425962  0.0007
n            -2.92048  1.413263  14 -2.066478  0.0578
TLS          60.36728 25.201710  14  2.395364  0.0311

Number of Observations: 329
Number of Groups: 17
```

Figure 11 Output from R computer package for mean rank on Burmese test

The key points to note from this are illustrated by the graph in Figure 12.

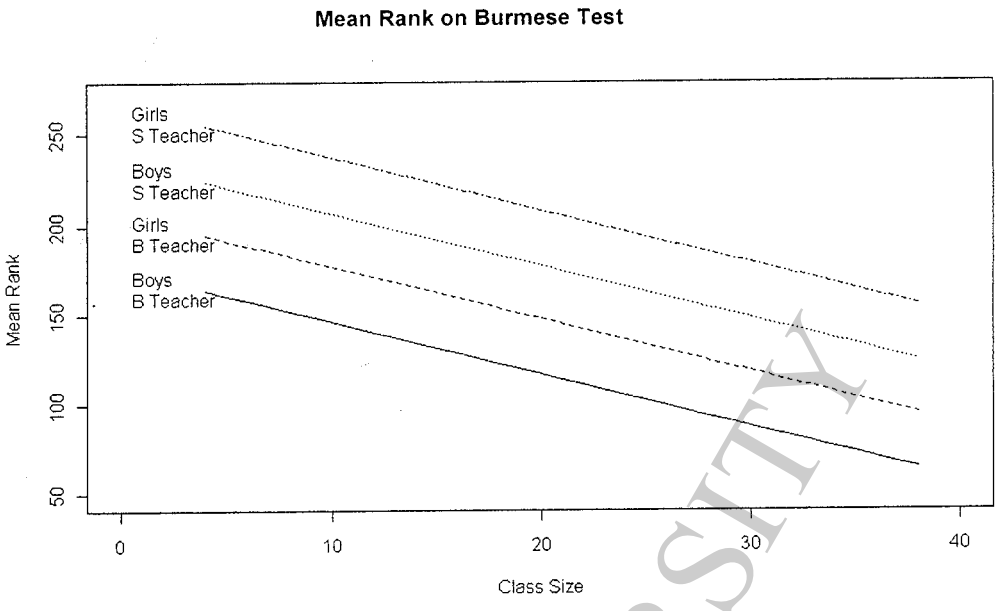


Figure 12 Mean rank on Burmese test

Figure 12 shows four lines representing the predictions of the model for the average (mean) class rank, one for each Gender-Teacher Language combination. A high mean rank represents high overall performance in a class. There are 329 learners altogether, so the ranks of individual student scores in Burmese have a range of 1 to 329. Each line decreases as class size increases implying that learners in smaller classes on average get higher scores than learners in larger classes. Gender is shown to have an effect because the line representing female learners is above the line for male learners for both Sgaw teachers and Burmese teachers. Teacher language is shown to have an effect because both lines for Sgaw teachers are above the two lines for Burmese teachers. The coefficient for Teacher Language in the model is 60.37 (2 d.p.) which means that for a given class size, learners whether male or female would on average have their rank improved by 60 if they had a Sgaw teacher instead of a Burmese teacher.

For ease of interpretation, the model is fitted to the actual Burmese test scores. The relevant output from the R computer package is shown in Figure 13.

```
Linear mixed-effects model fit by REML

Random effects:
 Formula: ~1 | class
        (Intercept) Residual
StdDev:     8.75503 16.73723

Fixed effects: Burmese ~ gender + n + TL
               Value Std.Error  DF   t-value p-value
(Intercept) 71.30422  6.332844 311 11.259431  0.0000
gender       6.54068  1.885984 311  3.468046  0.0006
n           -0.61428  0.295882  14 -2.076094  0.0568
TLS         12.85741  5.276429  14  2.436764  0.0288

Number of Observations: 329
Number of Groups: 17
```

Figure 13 Output from R computer package for mean score on Burmese test

Figure 14 helps to illustrate the key features of the fitted model.

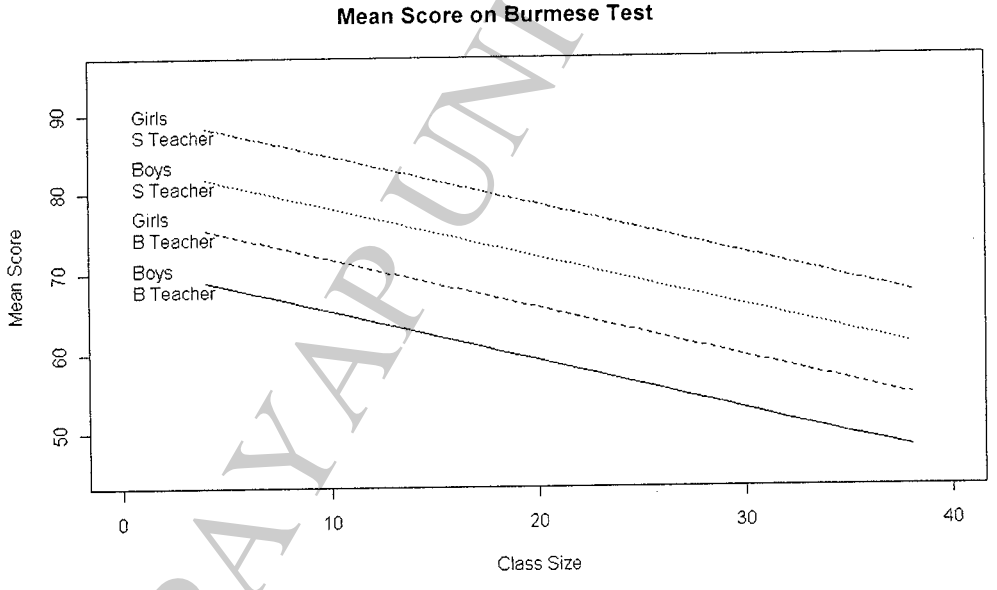


Figure 14 Mean score on Burmese test

Figure 14 shows four lines representing the predictions of the model for the average (mean) class score, one for each Gender-Teacher Language combination. Each line

decreases as class size increases implying that learners in smaller classes on average get higher scores than learners in larger classes. Gender is shown to have an effect because the line representing female learners is above the line for male learners for both Sgaw teachers and Burmese teachers. Teacher language is shown to have an effect because both lines for Sgaw teachers are above the two lines for Burmese teachers. The coefficient for Teacher Language in the model is 12.86 (2 d.p.) which means that for a given class size, a typical student whether male or female would have their score improved by approximately 13% if they had a Sgaw teacher instead of a Burmese teacher. This improvement is large enough to be of practical importance.

5.4.2 Analysis of Science test scores

The best model for ranked Science scores turned out to include interaction terms as well as individual factors. Figure 15 shows relevant output from the R computer package.

```
Linear mixed-effects model fit by REML

Random effects:
 Formula: ~1 | class
        (Intercept) Residual
StdDev:    41.77581 80.42206

Fixed effects: srank ~ gender + n + TL + gender:n + gender:TL + n:TL
+ gender:n:TL
                 Value Std.Error  DF    t-value p-value
(Intercept)   142.38716  57.99760 308  2.4550527  0.0146
gender        167.99627  55.55933 308  3.0237276  0.0027
n              -0.35354   3.30067  13 -0.1071120  0.9163
TLS           105.74594  71.57372  13  1.4774409  0.1634
gender:n       -6.59752   2.87871 308 -2.2918351  0.0226
gender:TLS   -201.12543  66.63374 308 -3.0183723  0.0028
n:TLS          -3.49487   3.71476  13 -0.9408074  0.3640
gender:n:TLS    9.47231   3.19594 308  2.9638571  0.0033

Number of Observations: 329
Number of Groups: 17
```

Figure 15 Output from R computer package for mean rank on Science test

The data in Figure 15 shows that the three-way interaction term gender:n:TLS is significant and is needed in the model. This implies that all of the other terms are also needed. In other words, Gender, Class Size, and Teacher Language along with the random class effect are needed in the model that best predicts rank of science scores. However, the relationship is complex because the effect of varying one factor depends on the levels of the other factors. Figure 16 shows the model predictions for mean ranks of science.

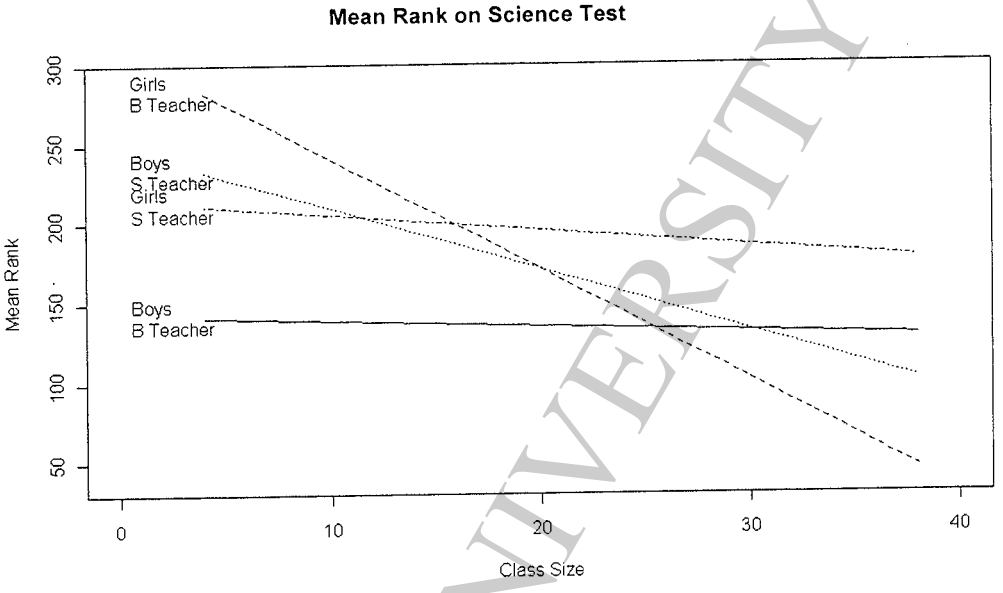


Figure 16 Mean rank on Science test

Figure 16 shows that for each Teacher Language-Gender combination, mean rank decreases with class size. The effect of increasing class size is approximately the same for Females with a Sgaw teacher and Males with a Burmese teacher: both show a slight decrease in mean rank as class size increases. However, increasing class size affects Girls with a Burmese teacher and Boys with a Sgaw teacher more acutely. For the range of class sizes the following observations can be made:

For Girls in classes with less than 16 learners, having a Burmese teacher leads to higher performance, whereas for classes larger than 16 having a Sgaw teacher leads to higher performance. The mean rank for girls with a Burmese teacher shows a dramatic decline as class size increases. For boys in classes with less than 30 learners having a Sgaw teacher leads to higher performance, whereas for classes larger than 30 having a Burmese teacher leads to higher performance.

5.4.3 Analysis of Mathematics test scores

The data in Figure 17 is the relevant output from the R computer package.

```
Linear mixed-effects model fit by REML

Random effects:
 Formula: ~1 | class
        (Intercept) Residual
StdDev:    64.54845 75.79769

Fixed effects: mrank ~ gender + n + TL + gender:n + gender:TL + n:TL
+ gender:n:TL
                  Value Std.Error  DF    t-value p-value
(Intercept)   119.44582  79.47835 308  1.5028725  0.1339
gender        142.13487  52.43433 308  2.7107216  0.0071
n               1.73016   4.60608  13  0.3756257  0.7133
TLS            56.52478  96.91592  13  0.5832352  0.5697
gender:n       -6.11498   2.71584 308 -2.2515955  0.0251
gender:TLS   -182.01964  63.03019 308 -2.8878167  0.0042
n:TLS          -2.45137   5.15790  13 -0.4752652  0.6425
gender:n:TLS    8.42364   3.01892 308  2.7902826  0.0056

Number of Observations: 329
Number of Groups: 17
```

Figure 17 Output from R computer package for mean rank on Mathematics test

The data in Figure 17 shows that the three-way interaction term gender:n:TLS is significant and is needed in the model. This implies that all of the other terms are also needed. In other words, Gender, Class Size, and Teacher Language along with the random class effect are needed in the model that best predicts Mathematics scores. However, the relationship is complex because the effect of varying one factor depends on the levels of the other factors. Figure 18 illustrates the key features of the model.

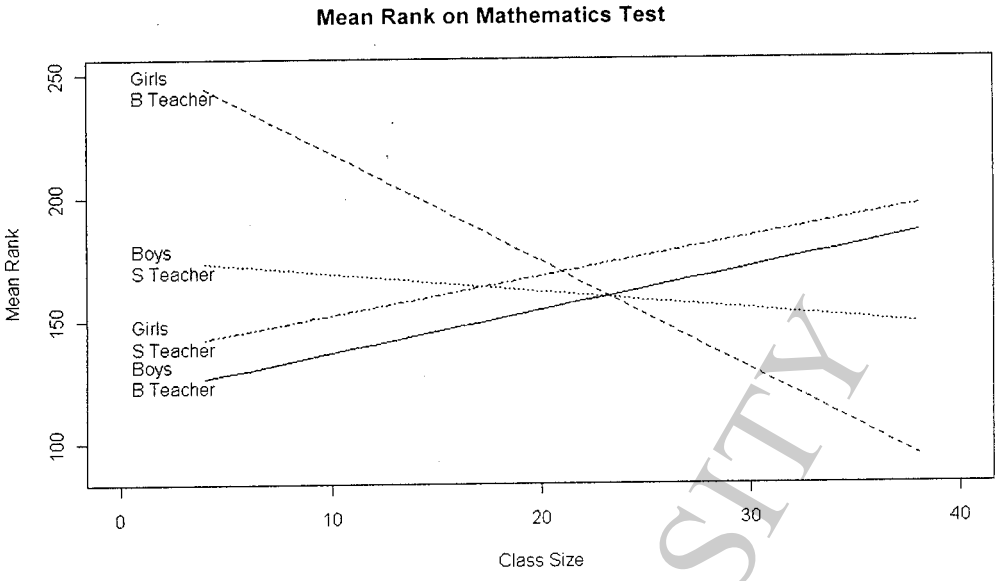


Figure 18 Mean rank on Mathematics test

Figure 18 shows that for both boys and girls class size has the opposite effect on classes with a Sgaw teacher to its effect on classes with a Burmese teacher. For classes smaller than 22, girls with Burmese teachers score higher whereas or classes larger than 22, girls with Sgaw teachers score higher. For boys these results are reversed: for classes smaller than 23, boys with Sgaw teachers score higher whereas for classes larger than 23, boys with a Burmese teacher score higher.

5.4.4 Analysis of English test scores

The best model for English test scores was based on a random effect for class plus the variable Gender. In other words it was better to omit Class Size and Teacher Language from the model. The data in Figure 19 is the relevant output from the R computer package.

```
Linear mixed-effects model fit by REML

Random effects:
 Formula: ~1 | class
        (Intercept) Residual
StdDev:    57.98148 73.44893

Fixed effects: erank ~ gender
                Value Std.Error  DF  t-value p-value
(Intercept) 162.44795 15.306416 311 10.61306  0.0000
gender       17.60931  8.289542 311  2.12428  0.0344

Number of Observations: 329
Number of Groups: 17
```

Figure 19 Output from R computer package for mean rank on English test

Figure 20 shows the plot of mean score versus class size illustrates the model.

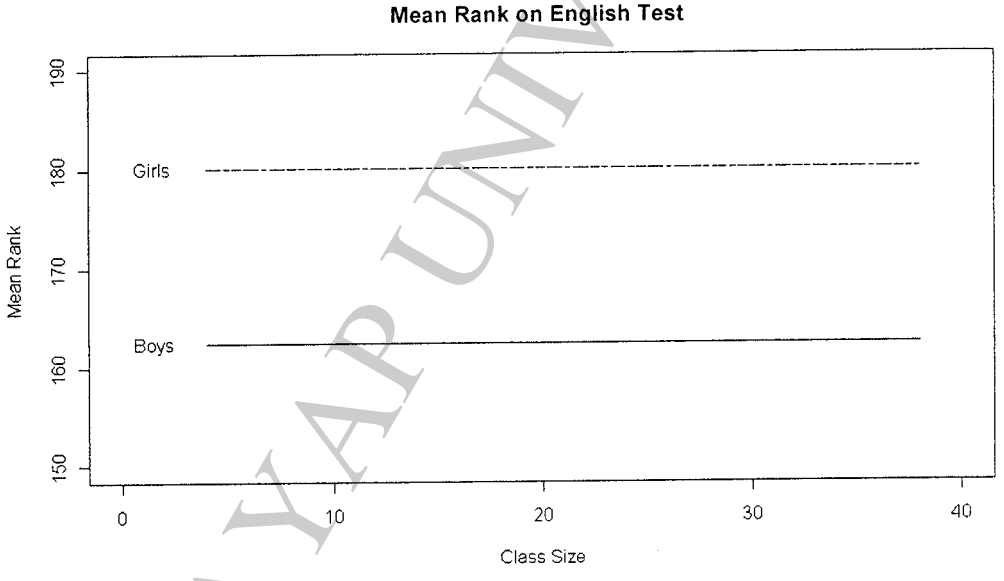


Figure 20 Mean rank on English test

The two lines are horizontal which shows that Class Size makes no difference to average (mean) class scores. Similarly Teacher Language makes no difference to mean test scores in the model – the only difference is due to Gender. The vertical

difference between the two lines is around 18, meaning that the typical rank of a Female student is 18 places higher than the typical rank of a Male student.

5.5 Summary of Findings

The findings revealed an important issue, gender is shown to be a vital factor in predicting student performance for each of the four academic subjects. In almost every case, female learners perform better than male learners.

A random class effect is shown to be vital for all four academic subjects. This captures something of the unmeasured factors such as teacher proficiency and 'environmental' factors particular to each class.

Grade 3 teacher language is shown to be vital for three of the four academic subjects. In almost every case, classes with a Sgaw-speaking teacher perform better than classes with a Burmese-speaking teacher.

Class Size is shown to be a vital factor for three of the four academic subjects. In almost every case, an increase in class size is associated with a decrease in predicted performance.

Although the sample only included one class that had not been taught by Sgaw teachers in Grades 1 and 2, the generally low scores of that class support the expectation that learning would be greatly impeded by not having a language in common with the teacher in these first two years in school. Even if learners have had Sgaw teachers in Grades 1 and 2, it is also generally detrimental to learning to not have a Sgaw teacher in Grade 3. The test results analyzed in the previous section were from November. That means it is about halfway through the June-February school year, so even in such a short period of time, the impact of language of instruction is detectable.

The strength of results varies across subjects, but in probably the most important subject, namely Burmese, there is very strong evidence that having a Sgaw teacher enables the learners to achieve higher scores. Moreover the size of the benefit is large enough to be of practical importance. These results are all the more remarkable since all the tests taken by the learners are in Burmese, so even if learners with a Sgaw teacher gain an advantage by having explanations in their mother tongue, they must still know enough Burmese to understand the questions and be able to answer them in Burmese. It is expected that if the whole dataset from the original tests were available (including the scores of those learners that failed

the original test) these analyses would show even larger differences between the classes with Sgaw teachers and those with Burmese teachers.

5.6 Discussion of Results

The analysis of test results in the previous sections shows different behavior for the different academic subjects. Here is a possible explanation of why the results show strongest evidence for the advantage of having the teacher speak the learners' mother tongue in Burmese followed by Mathematics and Science followed by English:

Sgaw teachers are fluent in Burmese, so they are able to teach it well as a subject. The questions on the test are in the language of the subject. The advantage of Sgaw is greater than for other subjects because there are fewer additional skills to master to perform well on the test.

In Science and Mathematics, Sgaw teachers produce better results, but the fact that the learners have to learn the Burmese vocabulary as well as understand the concepts reduces their advantage in the tests. Often Sgaw teachers have not got a bachelor's degree so their lower level of education probably means they are less proficient at teaching Mathematics and Science. In Science and Mathematics Sgaw teachers produce better overall average results than Burmese teachers. This can be seen by subdividing the schools in Table 2 by Grade 3 teacher language and calculating the median class scores for each subject. These median class scores are presented in Table 3 below.

Table 3 Teacher language median results for the class data in Table 2

Grade 3 Teacher Language	Class Size	Male	Female	Burmese	Mathematics	Science	English
Burmese	20.5	11	7	56.25	61	64.5	73.5
Sgaw	24	11	12	82	73	80	70

Despite the apparent advantage of Sgaw over Burmese as shown by the greater overall averages achieved in Science and Mathematics, the unclear patterns in Figures 16 and 18 shows that the situation is complex. It is most likely that the advantage gained by Sgaw teachers being able to communicate with the students in their class is reduced by the fact that learners have to understand the Burmese vocabulary as well as the concepts. Also, Sgaw teachers often do not have a

bachelor's degree so their lower level of education probably means they are less proficient at teaching Science and Mathematics.

Sgaw teachers learn English less well than Burmese teachers. Their lower proficiency means they are not so good at teaching English as the Burmese teachers. This means that the advantage gained by being able to explain in Sgaw is reduced by the teachers' lower proficiency in the subject.

In addition to the suggested explanation given above for why the performance over different subjects is mixed, it is important to compare this study with previous research to put the data analysis and results in perspective. The following points are relevant:

Studies of MTB MLE normally compare performance over several years: this study was done where the difference in teacher language was over approximately six months.

Other studies of MTB MLE examine the use of the mother tongue not only as an oral language of instruction but also as a true language of instruction which included curriculum and materials in the learners' L1 as well. In this study, the mother tongue was used as an informal oral language of instruction, but the curriculum and all materials were in the National language. Moreover, the tests which generated the data for this analysis were in the National language.

Other MTB MLE studies have been conducted with greater consistency of qualifications among teachers. In this study, although no information was gathered on the qualifications of teachers, it is almost certain that there were considerable differences between qualifications between Burmese and Sgaw teachers.