A PUBLICATION OF THE CENTER FOR EDUCATIONAL MEASUREMENT
APRIL 1996 • VOLUME 7

PHILIPPINE JOURNAL OF EDUCATIONAL MEASUREMENT

Center for Educational Measurement

PHILIPPINE JOURNAL of EDUCATIONAL MEASUREMENT

APRIL 1996 • **VOLUME 7**

The Philippine Journal of Educational Measurement publishes empirical papers and nonempirical reports such as theoretical notes, specialized reviews, and commentary on issues related to educational testing, measurement, assessment, evaluation, and research in all of its many aspects.

Address all inquiries and submit manuscripts for publication to:

The Editor
Philippine Journal of Educational Measurement
Center for Educational Measurement
6/F Wing B, Concorde Condominium
Cor. Salcedo and Benavidez Streets
Legaspi Village, 1229 Makati City

Articles and commentary in the Journal represent the views of the authors and do not necessarily reflect official CEM policy.

TABLE OF CONTENTS

DEVELOPMENT AND VALIDATION OF A READABILITY FORMULA FOR PHYSICS INSTRUCTIONAL MATERIALS

Jonathan A. Bayogan
Benguet State University

Abstract

The present study is an attempt to develop a readability formula for physics instructional materials. In Part 1, 21 teachers were asked to assign readability levels to 30 reading passages that were selected from books used in Philippine schools. Each passage was assigned a readability level which ranged form 4 to 16, representing the primary grades up to advanced or graduate level. Twelve variables were measured based on the textual and graphical characteristics of the passages. Factor analysis converged the variables into three factors: Quantitative Relationship and Sentence Structure (Factor 1), Technical Vocabulary (Factor 2), and Affective Scores (Factor 3). Multiple regression analysis using readability level as criterion resulted into three equally plausible readability formulas. In part 2, the developed readability formula was validated against existing readability formulas (Flesch formula, Fry formula, and Feedback-based communication index). Correlation analysis showed that all four readability formulas measured readability very similarly.

In recent years, published curriculum materials for elementary and secondary students started to increase. Textbooks, workbooks, and reference materials in different subject areas including the sciences were written in various series by several groups of authors.

This development indicates our growing independence from foreign books. It also marks a shift to curriculum materials developed by and for Filipinos. As this trend continues, it is relevant to develop appropriate instruments to determine the suitability of these materials to the intended audience. For example, which material matches what type of pupils, classes or grade level? How can a teacher or a school administrator be guided in choosing books for school use? What hints can one consider to rewrite books for better readability? These questions can be clarified when one has an instrument to measure the readability of instructional materials.

Research in readability was pursued using three general methodologies. They were: 1) survey of expert and reader opinion, 2) quantitative associational studies, and 3) experimental studies involving reader participation or feedback.

Readability indicators that were considered important by publishers, librarians, and teachers in judging the readability of a book included content, style, format, and organization (Gray and Leary, 1935 as cited by Chall, 1958). Strang (as cited by Chall, 1958) reported that high school and college students favored the stylistic features of the book such as "plain everyday English", "easy", "simple vocabulary", and "short paragraphs and sentences". A recent study by Spiegel and Wright (1984) showed Biology teachers' preference in textbook characteristics like graphics quality, reference to practical and real-life situations, inclusion of recent findings, and textbook features.

Quantitative associational findings revealed that average sentence length, average word length, personal words and sentences, word difficulty index, number of simple sentences, number of prepositional phrases and vocabulary factors, number of different technical words, number of different hard non-technical words, number of indeterminate clauses, and number of personal pronouns correlated significantly with the readability criterion (Flesch, 1943 as cited by Chall, 1958; Ojemann, 1934 as cited by Chall, 1958; Dale and Tyler, 1934 as cited by Chall, 1958; Gray and Leary, 1935 as cited by Chall, 1958).

Factors such as content of passages, writer's style, readers' background and readers' ability emerged as important variables in experimental studies (Grouws and Robinson, 1973).

Various readability tests and readability formulas were derived from the combinations of the above-mentioned indicators. Some of these formulas include *Flesch's Formula, Fry Readability Graph, and Dale-Chall Formula,* among others.

In the Philippines, some researchers investigated readability using these formulas. The Flesch formula was used by Espartero (1976), de la Cruz (1966), Lagarde (1984), Young (1991), and Cañares (1991). The Fry procedure was applied by Perez (1982), Talisayon (1983) and Young (1991). The Dale-Chall method was used by Espartero (1976) and Talisayon (1983). Some interesting findings of these studies were those by de la Cruz who used the Flesch Formula and found that majority of elementary science books were about two grade levels higher than that of the intended audience. Perez found four out of six elementary science textbooks that matched the grade levels of intended users using the Fry graph. Espartero compared the reading difficulty of 12 secondary science textbooks and recommended the consideration of readability ratings as a basis for textbook selection.

The observed discrepancy in these findings as in other local studies (Lagarde, 1984; Galitano, 1974; Cañares, 1991; and Young, 1991) was suggested to be confounded by the language considered in the development of these formulas. The above-mentioned formulas were all developed using popular English language (Cañares, 1991) and popular reading materials as sample. Thus, a formula developed and validated for technical materials and content areas like physics may prove more appropriate for science materials. Also, adjustments for readability levels had to be done for Filipinos since the above formulas used readability levels that were based on English as a first language. Readability level standards that consider English as a second language may be more helpful for local purposes.

In addition, these readability tests were limited to the use of a word list (Dale-Chall) and variables such as word length and sentence length (Flesch, 1948 as cited by Chall, 1958 and Fry, 1968, 1977) as indicators of readability. Some scholars claim that qualitative variables such as syntax, complexity of ideas, cohesiveness of discussion, reinforcement through restatement and repetition, writing style, and student interest and motivation are neglected (McConnell, 1982).

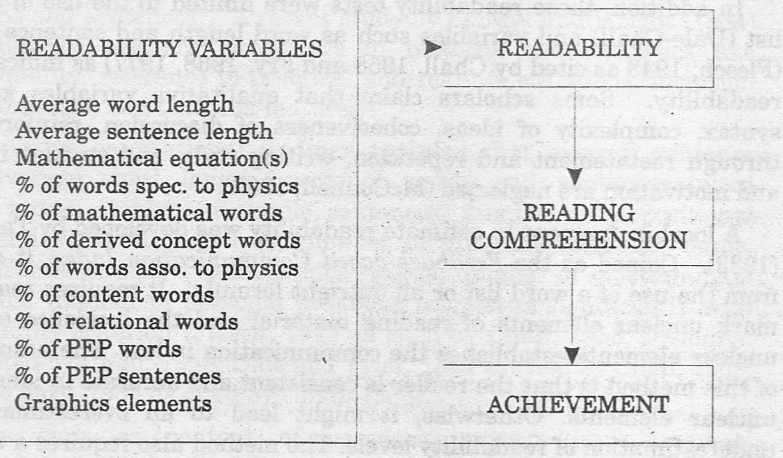
A local instrument to estimate readability was developed by Talisayon (1983). Coined as the *Feedback-based Communication Index*, it departs from the use of a word list or an outright formula. It requires readers to mark unclear elements of reading material and the incidence of these unclear elements establishes the communication index. The assumption of this method is that the reader is consistent and accurate in identifying unclear elements. Otherwise, it might lead to an overestimation or underestimation of readability levels. The method also requires a number of readers to establish a valid measure.

The present study is an attempt to develop an instrument to estimate the readability of physics instructional materials based on Filipino reader standards and on a content area using technical English specifically physics. It also sought to determine measurable variables which significantly contribute to the prediction of the readability of physics instructional materials. Specifically, the study attempted to answer the following questions:

1. What are the underlying factors related to the readability of physics instructional materials?
2. What readability variables significantly predict the readability level of physics instructional materials?
3. Is there a significant relationship between the readability levels predicted by the developed formula with those predicted by a) Flesch Formula b) Fry Formula and c) Feedback-based Communication Index?
4. Is there a significant relationship between the readability levels predicted by the developed formula with readability levels judged by readers?
5. Is the reading comprehension achievement of students significantly related to the readability of physics reading passages?

The present study also aims to develop a readability formula that would contribute towards a successful individualized instruction that requires readable books. The specific goal of this study is to use this formula as a good measure for evaluating curriculum materials for improvement. It is also hoped to provide assistance in deciding the compatibility of reading books to specific groups of readers, in estimating the reading difficulty of books, for grading books for lists, for predicting characteristics of writers and for writing or rewriting books to desired levels of difficulty. Therefore, the results of this study can benefit

Figure 1. Conceptual Paradigm



teachers, administrators, curriculum writers, curriculum planners, and educators specially those in the fields of science and mathematics.

CONCEPTUAL FRAMEWORK

Relevant findings on the nature of readability reviewed in this study laid down the foundation for its conceptual framework. The general thesis that a more readable instructional material improves reading comprehension, which in turn, influences student achievement sets the goal for selecting the variables with the most predictive ability in determining the level of difficulty of a reading material that is also amenable to reliable and objective quantitative measurement.

After careful deliberation, 12 readability variables were included (see Figure 1). These variables are defined below:

1. *Presence/absence of mathematical equation* (MEQ). This refers to the presence or absence of mathematical equation(s) in a passage. As a dummy variable, *1* represents presence and *0* represents absence of mathematical equation(s).

2. *Graphics elements contribution* (GRA). This refers to the presence and absence of graphics elements in the passages. It is also measured as a dummy variable.

3. *Average sentence length* (ASL). This is the total number of syllables divided by the total number of sentences in the passage.

4. *Average word length* (AWL). This is obtained by dividing the total number of syllables by the total number of words in the passage.

5. *Percentage of mathematical words* (PMW). This is obtained by dividing the total number of mathematical words by the total number of words in the passage times 100.

6. *Percentage of derived concept words* (PDCW). This is taken by dividing the total number of derived concept words by the total number of words in the passage times 100.
7. *Percentage of content words* (PCW). This is taken by dividing the total number of content words by the total number of words in the passage times 100.
8. *Percentage of relational words* (PRW). This is taken by dividing the total number of relational words by the total number of words in the passage times 100.
9. *Percentage of words specific to physics* (PWSP). This is taken by dividing the total number of words specific to physics by the total number of words in the passage times 100.
10. *Percentage of words associated to physics* (PWAP). This is taken by dividing the total number of words associated to physics by the total number of words in the passage times 100.
11. *Percentage of PEP words* (PPEPW). This is taken by dividing the total number of personal, events, and places words by the total number of words in the passage times 100.
12. *Percentage of PEP sentences* (PPEPS). This is taken by dividing the total number of personal, events, and places sentences by the total number of sentences in the passage times 100.

Furthermore, the above key measures were based on the following key terms:

Syllable. This is defined as a phonetic syllable. Generally there are as many syllables as vowel sounds. For example, *stopped* is one syllable and *wanted* is two syllables. Numerals and initialization are given one syllable for each symbol. Thus, *1994* is four syllables, *%* is one syllable and *CEM* is three syllables.

Word. This is defined as a group of symbols with a space on either side.

Content words. Words that carry ideas such as nouns, verbs, adjectives, adverbs and the like.

Relational words. Words which are considered important to higher level thinking; words that convey relationships between one concept to another as they are used in a sentence.

Derived concept words. Words that are a combination of fundamental concepts. For example, speed is a concept that combines distance and time.

Mathematical words. Words that imply mathematical operation, relationship or description (e.g., addition, ratio, square root, circle etc.).

Words specific to physics. Words marked by specific scientific (physics) interpretation.

Words associated to physics. This includes all mathematical words, prepositional words, relational terms and comparative words used to describe, relate, associate and make sense to physics term.

PEP word. Words that include persons, events and places. Persons include personal words like all pronouns in the first, second, and third persons whether singular or plural, nominative, objective and possessive; all words having masculine or feminine gender whether singular or plural; and all collective nouns (Flesch, 1948 as cited by Chall, 1958). It also includes names of persons, events, and places.

PEP sentence. This refers to spoken sentences including quoted sentences and "he said" statements. It also includes statements, questions, commands, and requests directly addressed to the reader (Flesch, 1948 as cited by Chall, 1958). As an extension, statements carrying names of persons, events, and places were included (Chall, 1958)

Graphics elements. These are non-textual elements such as graphs, photographs, cartoons, illustrations, figures, diagrams and the like.

METHODOLOGY

The first part is the development of the readability formula and the second part is the validation of the formula.

Part 1 : *Development of the Readability Formula*

Respondents. The first set of respondents was composed of 18 Science and 3 English teachers whose task was to judge the test items (in the form of passages) according to their suitable readability level.

Instrument. Thirty passages with no less than 100 words each were carefully chosen from eighteen sources consisting of 17 books and one lecture manual. Four of these sources were used in elementary, six in high school , and eight in college. Fourteen books were written by Filipino authors and four by foreign authors.

The passages were chosen according to the general topics commonly taken in the three levels (i.e., acceleration, force and gravitation, work and energy, and waves) and the differing degrees of difficulty based on the grade/year level where the passage was taken.

In judging the readability level, the following scale was used:

Grade Level	Readability Level
Primary Grades, Elementary	4
Intermediate Grades, Elementary	6
First & Second Year, High School	8
Third & Fourth Year, High School	10
First & Second Year, College	12
Third & Fourth Year, College	14
Advanced/Graduate Level	16

The readability levels correspond to the grade and year levels in the Philippine educational system. The lowest level (Level 4) and the highest level (Level 16) correspond to the primary grades and the graduate level, respectively. For example, if the judges decide that a given passage was easily understood by primary grade pupils, the readability level of the material is Level 4 and so forth.

Analysis of Data. Before the raw scores were transmuted into various statistical equivalents, the three highest scores and the three lowest scores from the 21 judged readability level entries were excluded to reduce bias and variability. Thus, only 15 entries were used to compute the mean readability level for each passage. The computed means were considered as the readability level for the corresponding passages and subsequently used as the criterion variable.

Cronbach's alpha was computed three times to determine the reliability of the judges' responses: (1) in general, (2) according to the levels they taught (elementary, high school, college), and (3) according to their teaching fields (English or Science). The resulting magnitudes were high at 0.98, 0.98, and 0.94, respectively.

For the analysis of data, the independent variables (i.e., the 12 readability variables) were factor analyzed using principal component analysis and varimax rotation. This is to address the first hypothesis of the study which states that readability is multi-dimensional, not unitary.

The readability formula was developed using the multiple regression analysis in two ways: (1) by using one surrogate variable each from factors found in the factor analytic procedure and (2) by including all independent variables in the analysis.

Each word type/sentence type that appeared in the passage corresponds to one count regardless of repetitions. This means that if the word "energy" appeared 10 times in the reading passage, it was given a word count of 10.

RESULTS

Judged Readability Level

As a whole, the judges generally agreed on the readability level of each of the reading passages as shown in Tables 1 and 2.

The judged readability level ranged from 5.06 to 13.73, spanning from the intermediate level in elementary (Levels 5 to 6) up to the extreme end of the collegiate level (Levels 13 to 14). The frequency of the passages' judged readability levels and the relatively good fit on the normal curve indicates that, with a big sample size, the readability levels of reading passages follow the normal distribution.

Factor Analysis

Initially, it was necessary to test the appropriateness of factor analysis. Thus the correlation matrix of the independent variables and

Table 1
Mean, Standard Deviation, Minimum and Maximum Judged Readability Levels (READ) of 30 Passages

Passage Number	Mean READ	SD	High	Low	Passage Number	Mean READ	SD	High	Low
1	10.53	0.92	12	10	16	13.60	0.83	14	12
2	8.80	1.01	10	8	17	11.73	1.03	14	10
3	11.20	1.01	12	10	18	11.73	1.49	14	10
4	8.40	1.35	10	6	19	9.33	0.96	10	8
5	7.87	1.60	10	6	20	5.20	1.47	8	4
6	9.33	1.45	12	8	21	10.13	0.52	12	10
7	10.00	0.00	10	10	22	6.40	1.35	10	4
8	5.06	1.49	8	4	23	11.06	1.03	12	10
9	5.87	1.41	8	4	24	10.13	0.52	12	10
10	7.87	1.19	10	6	25	9.06	1.03	10	8
11	10.00	0.00	10	10	26	11.87	1.18	14	10
12	9.33	1.23	12	8	27	6.93	1.28	8	4
13	10.93	1.03	12	10	28	9.73	0.70	10	8
14	13.33	0.98	14	12	29	9.33	0.98	10	8
15	13.73	1.49	16	12	30	8.00	1.07	10	6

READ	Freq.	Valid Percent	Cum Percent	READ	Freq.	Valid Percent	Cum Percent
5.06	1	3.3	3.3	10.00	2	6.7	60.0
5.20	1	3.3	6.7	10.13	2	6.7	66.7
5.87	1	3.3	10.0	10.53	1	3.3	70.0
6.40	1	3.3	13.3	10.93	1	3.3	73.3
6.93	1	3.3	16.7	11.06	1	3.3	76.7
7.87	2	6.7	23.3	11.20	1	3.3	80.0
8.00	1	3.3	26.7	11.73	2	6.7	86.7
8.40	1	3.3	30.0	11.87	1	3.3	90.0
8.80	1	3.3	33.3	13.33	1	3.3	93.3
9.06	1	3.3	36.7	13.60	1	3.3	96.7
9.33	4	13.3	50.0	13.73	1	3.3	100.0
9.73	1	3.3	53.3				

Mean: 9.55 Std Dev: 2.29 Minimum: 5.06 Maximum: 13.73

Table 2
Mean, Standard Deviation, Maximum and Minimum of 12 Independent Variables of Readability

Var	Mean	SD	Kurt.	Skewness	Min.	Max.
MEQ	0.37	0.49	-1.78	0.58	0.00	1.00
GRA	0.47	0.51	-2.13	0.14	0.00	1.00
ASL	27.49	14.05	1.31	1.13	7.71	66.67
AWL	1.51	0.21	-0.83	0.04	1.13	1.89
PMW	7.76	7.49	1.42	1.12	0.00	30.77
PDCW	7.01	6.09	1.22	1.09	0.00	24.29
PCW	59.94	7.18	3.43	1.67	51.22	84.07
PRW	4.70	3.18	-0.26	0.34	0.00	12.42
PWSP	11.66	6.85	-0.66	0.21	0.00	25.17
PWAP	13.20	7.89	0.48	0.60	1.68	35.16
PPEPW	2.59	2.95	0.04	1.12	0.00	9.24
PPEPS	24.90	23.08	-0.37	0.73	0.00	75.00

Number of Valid Observations (Listwise) = 30.00

Legend:

MEQ - presence/absence of mathematical equation
GRA - presence/absence of graphics elements
ASL - average sentence length
AWL - average word length
PMW - percentage of mathematical words
PDCW - percentage of derived concept words
PCW - percentage of content words
PRW - percentage of relational words
PWSP - percentage of words specific to physics
PWAP - percentage of words associated to physics
PPEPW - percentege of personal, events, and places words
PPEPS - percentage of personal, events, and places sentences

the readability level (READ) was computed (*see Table 3*). The matrix revealed that GRA, PCW, and PRW variables did not correlate significantly with any other variables. Six variables, namely, MEQ, ASL, AWL, PMW, PWSP, and PWAP related positively with the readability level. This meant that higher measures for each variable resulted to a higher readability level. Only one variable, PPEPW, related negatively with readability level.

Based on the magnitudes of correlation coefficients of the variables, the best single predictor of readability level was ASL (0.843), followed by PMW (0.726) and PWSP (0.654).

Table 4 shows that the Bartlett test of Sphericity exhibited a high magnitude of 217.95. This indicates that the population correlation matrix was unlikely an identity matrix and that the sample came from a multivariate normal population.

As an additional indicator of the strength of relationship among the variables, the Anti-Image Correlation Matrix whose value corresponds to a negative partial correlation coefficient, showed a large proportion of low coefficients. This means that the variables shared common factors since the partial correlation coefficient is the estimate of the correlations between unique factors. For a factor analysis to proceed, unique factors must be correlated with each other (Norusis, 1988).

The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was computed twice. The first computation using the 12 variables (*see Table 4*) obtained a value of 0.57 which bordered around the "miserable" level (Norusis,1988). To raise this level, variables with the smallest value in the anti-image correlation matrix (i.e., PCW and PRW) were excluded in the second computation. As shown in Table 5 the KMO-MSA was raised to 0.66 which was within the "mediocre" level and near the "middling" level which, nevertheless, merits the use of factor analysis.

The next step in factor analysis is the factor extraction using the principal components analysis. This will determine the number of factors necessary to represent the data. A criterion value of 60% for the total variance and 1.00 for the eigenvalue was set (Hair et. al, 1990). Actual analysis converged the variables into three factors that accounted for 71.7% of the total percent variance (*see Table 5*).

To minimize the number of variables that have high loading in a factor, the obtained factors were rotated using the varimax rotation as seen in Table 6. The criterion for factor loading was set at 0.5000.

The final tabulation showed four variables loading significantly in Factor 1, four in Factor 2, and two in Factor 3. The list of factors and variables are as follows:

Factor 1 : percentage of mathematical words (PMW), percentage of words associated to physics (PWAP), presence/ absence of mathematical equations (MEQ), and average sentence length (ASL)

Factor 2 : presence/absence of graphics elements (GRA), percentage

Table 3

Correlation Matrix of Independent Variables and Readability Level

	MEQ	GRA	ASL	AWL	PMW	PDCW	PCW	PRW	PWSP	PWAP	PPEPW	PPEPS	READ
MEQ	1.000	-.018	.600**	.260	.377	.187	-.073	.211	.273	.383	-.034	.207	.517*
GRA	-.018	1.000	-.275	-.353	-.057	-.335	-.085	.124	-.237	.066	.144	.084	-.143
ASL	.600**	-.275	1.000	.607**	.552**	.225	-.014	-.101	.542**	.412	-.396	-.206	.843**
AWL	.260	-.353	.607**	1.000	.536*	.304	.178	-.227	.504*	.362	-.326	-.181	.533*
PMW	.377	-.057	.552**	.536*	1.000	.394	.256	.041	.584**	.920**	-.363	-.189	.726**
PDCW	.187	-.335	.225	.304	.394	1.000	.008	-.075	.586**	.298	-.145	-.128	.262
PCW	-.073	-.085	-.014	.178	.256	.008	1.000	-.207	.087	.276	.126	.268	.097
PRW	.211	.124	-.101	-.227	.041	-.075	-.207	1.000	-.259	.351	-.075	-.018	-.045
PWSP	.273	-.237	.542**	.504*	.584**	.586**	.087	-.259	1.000	.394	-.359	-.332	.654**
PWAP	.383	.066	.412	.362	.920**	.298	.276	.351	.394	1.000	-.305	-.118	.606**
PPEPW	-.034	.144	-.396	-.326	-.363	-.145	.126	-.075	-.359	-.305	1.000	.858**	-.423*
PPEPS	.207	.084	-.206	-.181	-.189	-.128	.268	-.018	-.332	-.118	.858**	1.000	-.215
READ	.517*	-.143	.842**	.533*	.726**	.262	.097	-.044	.654**	.606**	-.423*	-.215	1.000

Number of cases: 30   1-tailed Signif: * - .01   ** - .001

Table 4

Barlett Test of Sphericity, Anti-Image Correlation Matrix and Kaiser-Meyer-Olkin Measure of Sampling Adequacy of 12 Readability Variables

Bartlett Test of Sphericity = 217.95
Significance = .00000

Anti-Image Correlation Matrix

	MEQ	GRA	ASL	AWL	PMW	PDCW	PCW	PRW	PWSP	PWAP	PPEPW	PPEPS
MEQ	.577											
GRA	-.185	.424										
ASL	-.545	.251	.750									
AWL	.074	.155	-.289	.874								
PMW	.028	.218	.040	.229	.570							
PDCW	-.113	.356	.269	-.030	.031	.656						
PCW	.158	.250	.139	-.122	.366	.182	.268					
PRW	.208	.308	.143	-.036	.732	.073	.469	.187				
PWSP	-.122	-.066	-.189	-.017	-.235	-.464	-.160	.033	.791			
PWAP	.028	-.320	-.024	.163	-.954	-.099	-.473	-.816	.165	.483		
PPEPW	.215	-.093	.123	.086	.014	-.037	.171	.023	-.154	.030	.621	
PPEPS	-.439	.093	.016	-.063	.011	-.006	-.303	.053	.278	-.036	.853	.499

Note: Measures of sampling adequacy (MSA) are printed on the diagonal. Kaiser-Meyer-Olkin Measure of Sampling Adequacy = 0.57.

Table 5
KMO Measure of Sampling Adequacy and Eigenvalues of 10 Readability Variables

FACTOR VARIABLES:
MEQ, GRA, ASL, AWL, PMW, PDCW, PWSP, PWAP, PPEPW, PPEPS

Final Statistics			
Factor	Eigenvalue	Pct. of Var	Cum Pct
1	4.124	41.20	41.20
2	1.731	17.30	58.50
3	1.313	13.10	71.70

Note: Kaiser-Meyer-Olkin Measure of Sampling Adequacy = .66

Table 6
Rotated Factor Matrix of 10 Variables

	Rotated Factor Matrix		
	Factor 1	Factor 2	Factor 3
MEQ	.68319 +	.16371	.31572
GRA	.18145	-.84664 +	.01905
ASL	.63837 +	.45665	-.14951
AWL	.45939	.57946 +	-.17851
PMW	.88229 +	.16625	-.22760
PDCW	.26560	.63978 +	-.04926
PWSP	.49131	.56639 +	-.29177
PWAP	.88005 +	-.04167	-.17019
PPEPW	-.22080	-.14220	.90029 +
PPEPS	.01023	-.09661	.96212 +

Note: Factor loading criterion: 0.5000; + - Variables included in the factor

of derived concept words (PDCW), average word length (AWL), and percentage of words specific to physics (PWSP)

Factor 3 : percentage of PEP words (PPEPW) and percentage of PEP sentences (PPEPS)

These results confirmed the hypothesis that the readability of physics instructional material is multi-dimensional and it is likely to have at least three dimensions. These are: **Quantitative Relationship and Sentence Structure** (Factor 1), **Technical Vocabulary** (Factor 2), and **Affective Score** (Factor 3).

Developing the Readability Formula

After identifying the factors of readability, we are now ready to develop the readability formula. The first step was to run a multiple linear regression model using all independent variables as predictors with readability level as criterion. The full model which included all the ten (10) independent variables as predictors (*see Table 7*) accounted for 85.03% of the total variance and the adjusted percent variance explained was 77.16% The latter was a preferred measure of goodness of fit because it was not subject to the inflationary bias of unadjusted R^2 (Norusis, 1988). However, this model does not lend itself to easy administration so the researcher opted for the next better choice, that is to consider the model which included only the surrogate variables.

Selection of surrogate variables was based on the magnitude of standardized beta weights (*see Table 7*) of each predictor variable. The surrogate variables with relatively large contributions and their corresponding factors were: ASL (0.620) for Factor 1 , PWSP (0.285) for Factor 2, and PPEPS (0.221) for Factor 3.

Table 7
Multiple Regression Model Using 10 Independent Variables as Predictors of Readability

	Variables in the Equation				
Variable	B	SE B	Beta	T	Sig T
PPEPS	.0217	.02048	.2208	1.061	.3019
GRA	.1915	.47669	.0427	.402	.6924
PWAP	.0122	.08096	.0422	.150	.8821
MEQ	-.2463	.62479	-.0531	-.394	.6978
PDCW	-.0398	.04474	-.1065	-.889	.3852
AWL	-1.2220	1.36222	-.1152	-.897	.3809
PWSP	.0946	.04933	.2851	1.917	.0704
ASL	.1003	.02611	.6203	3.841	.0011
PPEPW	-.1659	.16015	-.2154	-1.036	.3133
PMW	.0814	.09855	.2684	.826	.4190
Constant	6.8470	2.15555		3.176	.0050

Multiple R	.9221
R Square	.8503
Adjusted R Square	.7716
Standard Error	1.0858

Analysis of Variance

Sources of Variation	df	Sum of Squares	Mean Squares
Regression	10	127.2665	12.7266
Residual	19	22.3996	1.1789

F = 10.7951 Signif F = .0000

These surrogate variables accounted for 76.52% of the total variance while the adjusted proportion of variance explained was 73.81% using the multiple regression model (*see Table 8*) with the greatest contribution by average sentence length (ASL) at 0.6920 followed by percentage of words specific to physics (PWSP) at 0.2862. The contribution of PPEPS was negligible and insignificant.

The removal of the PPEPS variable in the backward regression analysis (*see lower half of Table 8*) accounted for practically the same amount of variance (76.47% for the total variance and 74.73% for the adjusted amount of variance) by ASL and PWSP variables with corresponding beta weights of 0.692 and 0.286, respectively. This model is the first plausible formula written as:

Equation 1: Readability Level Using Physics Words (RPL)

$$RLP = 5.336 + 0.112\ ASL + 0.093\ PWSP$$

Another option explored in developing a readability formula was to conduct a stepwise regression analysis of all the ten predictor variables regardless of what factors the variables were classified with. The results of this analysis is shown in Table 9.

The significant predictors with high standardized beta weights were ASL (0.635) and PMW (0.376). This equation which account for 80.81% of the total variance and 79.39% of the adjusted R^2 was substantially higher than the previous equation. This equation also had a more balanced distribution of standardized beta weights; ASL with 0.6350 and PMW with 0.3761. This meant that PMW contributed better in the regression model than did PSWP in the previous equation. This equation is written as:

Equation 2: Readability Level Using Mathematical Words (RLM)

$$RLM = 5.781 + 0.114\ PMW + 0.103\ ASL$$

The improvement of the regression model with PMW in the equation, instead of PWSP, hinted that both variables contributed well to the prediction of readability along with ASL.

Therefore, a compromise between the two formulas was considered by the researcher. A step-by-step regression analysis was conducted. Table 10 revealed that the combination of ASL, PMW and PWSP variables accounted for 82.18% of the total variance.

Using the computed coefficients, the third formula can now be written as:

Equation 3: Readability Level Using Mathematical and Physics Words (RLMP)

$$RLMP = 5.552 + 0.095\ ASL + 0.095\ PMW + 0.051\ PWSP$$

Comparison of the three models in terms of accounted total variance, residual statistics, residual variability (*see Table 11*), range of prediction,

Table 8
Multiple Regression Model Using Surrogate Variables as Predictors of Readbility

	Using All Surrogate Variables in the Equation				
Variables	B	SE B	Beta	T	Sig
PPEPS	.00222	.0099	.0225	.223	.8249
ASL	.11188	.0183	.6920	6.118	.0000
PWSP	.09493	.0389	.2862	2.438	.0219
Constant	5.24939	.6319		8.308	0.0000
	Multiple R		.8747		
	R Square		.7652		
	Adjusted R Square		.7381		
	Standard Error		1.1626		

	Using Significant Variables in the Equation				
Variables	B	SE B	Beta	T	Sig
ASL	.1118	.0180	.6912	6.224	0.0000
PWSP	.0926	.0368	.2791	2.513	.0182
Constant	5.3355	.4915		10.848	0.0000
	Multiple R		.8745		
	R Square		.7647		
	Adjusted R Square		.7473		
	Standard Error		1.1420		

Table 9
Final Result of Stepwise Multiple Regression Using All Independent Variables as Predictors of Readability

Variables	B	SE B	Beta	T	Sig
ASL	0.1027	0.0163	0.6350	6.283	.0000
PMW	0.1140	0.0306	0.3761	3.722	.0009
Constant	5.7805	0.4194		13.782	.0000
	Multiple R		.8990		
	R Square		.8081		
	Adjusted R Square		.7939		
	Standard Error		1.0313		

	Analysis of Variance		
Sources of Variation	df	Sum of Squares	Mean Square
Regression	2	120.9477	60.4738
Residual	27	28.7185	1.0636
F = 56.8552		Signif F = .0000	

Table 10
Multiple Regression Model Using PWSP, PMW, and ASL as Predictors of Readability

Variables	B	SE B	Beta	T	Sig T
PWSP	.0506	.0358	.1524	1.414	.1693
ASL	.0949	.0169	.5869	5.594	.0000
PMW	.0951	.0330	.3136	2.886	.0077
Constant	5.5516	.4426		12.544	.0000
	Multiple R		.9065		
	R Square		.8218		
	Adjusted R Square		.8013		
	Standard Error		1.0128		

Analysis of Variance

Sources of Variation	df	Sum of Squares	Mean Square
Regression	3	122.9975	40.9992
Residual	26	26.6687	1.0257
F = 39.9711		Signif F = .0000	

Table 11
Comparative Statistics of the Three Readability Formula

Statistics	RLMP			RLM			RLP		
Multiple R	.9065			.8990			.8745		
R^2 Value	.8218			.8081			.7647		
Adjusted R^2	.8013			.7939			.7473		
Residuals	Min	Max	SD	Min	Max	SD	Min	Max	SD
Predicted	6.55	13.86	2.06	6.57	13.77	2.04	6.65	14.83	1.99
Residual	-1.84	-1.84	0.96	-1.9	1.91	0.99	-2.20	1.92	1.10

standardized residual, standardized predictions, outliers (*see Table 12*), and other significant criteria revealed that the best choice for readability formula was the third formula or the RLMP.

This formula, therefore, is renamed JB Formula (JB is the author's initials). In the formula, the criterion is also renamed RDS for Reading Difficulty Score. This is so because higher scores indicate difficult passages and lower scores indicate easier passages. This formula is written as:

Equation 4: JB Formula

$$RDS = 5.5516 + 0.0949\ ASL + 0.0951\ PMW + 0.0506\ PWSP$$

Table 12

Outliers and Histogram of Standard Residuals of Three Readability Equations

Outliers of Standardized Residual

RLMP or JBRF		RLP		RLM	
Case	ZRESID	Case	ZRESID	Case	ZRESID
9	-1.8146	9	-1.9276	9	-1.8602
17	1.8125	8	-1.8353	17	1.8553
8	-1.7788	25	1.6868	20	-1.7304
25	1.4947	17	1.6052	24	-1.6475
20	-1.4461	5	-1.5897	8	-1.6273
24	-1.4380	21	1.4644	19	1.4906
19	1.3016	20	-1.2665	13	1.3282
21	1.1304	16	-1.0805	18	-1.0581
5	-.9650	13	1.0556	6	1.0410
27	-.8720	11	.9289	25	1.0341

Histogram - Standardized Residual
(* = 1 Case, : = Normal Curve)

RLMP or JBRF			RLP			RLM		
N	Exp N		N	Exp N		N	Exp N	
0	.02	Out	0	.02	Out	0	.02	Out
0	.05	3.00	0	.05	3.00	0	.05	3.00
0	.12	2.67	0	.12	2.67	0	.12	2.67
0	.27	2.33	0	.27	2.33	0	.27	2.33
0	.55	2.00 .	0	.55	2.00 .	1	.55	2.00 :
1	1.00	1.67 :	2	1.00	1.67 :*	0	1.00	1.67 .
2	1.65	1.33 *:	1	1.65	1.33 *.	2	1.65	1.33 *:
3	2.42	1.00 *:*	3	2.42	1.00 *:*	5	2.42	1.00 *:***
4	3.19	-1.67 **:*	4	3.19	.67 **:*	0	3.19	.67 .
2	3.76	.33 **.	4	3.76	.33 ***:	5	3.76	.33 ***:*
6	3.97	.00 ***:**	3	3.97	.00 ***.	4	3.97	.00 ***:
4	3.76	-.33 ***:	4	3.76	-.33 ***:	6	3.76	-.33 ***:**
1	3.19	-.67 *	3	3.19	-1.67 **:	2	3.19	-1.67 **.
3	2.42	-1.00 *:*	2	2.42	-1.00 *:	1	2.42	-1.00 *.
2	-1.65	-1.33 *:	1	-1.65	-1.33 *.	0	-1.65	-1.33 .
2	1.00	-.67 :*	1	1.00	-.67 :	3	1.00	-.67 :**
0	.55	-2.00 .	2	.55	-2.00 :*	1	.55	-2.00 :
0	.27	-2.33	0	.27	-2.33	0	.27	-2.33
0	.12	-2.67	0	.12	-2.67	0	.12	-2.67
0	.05	-3.00	0	.05	-3.00	0	.05	-3.00
0	.02	Out	0	.02	Out	0	.02	Out

Part 2 : *Validation of the Readability Formula*

The JB formula was validated in three ways:

a) Comparisons of the predictive ability of the JB formula and the predictive abilities of Flesch readability formula, the Fry readability graph and Feedback-based Communication Index formula.

Flesch readability formula. This test includes two types of scores: the Reading Ease Score (RES) and the Human Interest Score (HIS). The Reading Ease Score which is based on the average word length and average sentence length was used for validation.

Fry readability graph. It is a mathematical relationship between the number of sentences and the number of syllables in a 100-word sample. These variables are plotted in the Fry graph to determine the readability level.

Feedback-based Communication Index formula. This accounts for the clarity of the elements of a reading material as perceived by readers. The incidence of unclear elements is the basis for the communication index.

b) Correlation analysis of the judged readability levels of the five passages and the readability levels as predicted by the JB formula.

c) Correlation analysis of the 25-item scores from the five passages and the readability levels as predicted by the JB formula.

Respondents. There were two sets of respondents in this experiment. First was a group of 67 college freshmen from two universities: one in the province and one in Metro Manila. Students from the Manila university are likely to speak Filipino as their first language while those from the provincial university are most likely to speak their native language. In calculating the Feedback-based communication index, these students were asked to read the passages very carefully. While reading, the students were asked to mark unclear elements in the passages. Elements could be a word, a phrase, a sentence, or a paragraph. Students were reminded every five minutes to mark unclear elements.

The same group of respondents was also asked to rank the passages according to the order of their perceived readability (1 to 5, from easiest to most difficult). The data were used in computing the judged readability level of the passages.

The second group of respondents was composed of 143 college freshmen from the same universities. Their task is to read the five passages and answer the 25-item reading comprehension test (5 multiple choice). The scores were used in the third method of validation.

Both groups of respondents, at the time of the study, were taking a physics course. Their major courses varied from physics, mathematics, chemistry, biology, agriculture, and education.

Instrument. Five passages were carefully selected from books used in elementary and college levels. The passages were selected so that the range of readability levels was wide based on the researcher's judgment and the group of audience for which the materials were intended for. One passage was taken from a textbook intended for Grade II pupils. Another one was taken from a college physics book with a conversational writing style. A sample item in the Scholastic Aptitude test (Practice Set) was also included and the rest were taken from a reference book for college students.

The readability levels of these passages were determined using the Flesch, Fry, Feedback-Based Communication Index and the JB formula. The readability level of each passage was also judged by student readers. Finally, the scores of students were taken from a 25-multiple-choice-item test prepared from the passages. The reliability of the test using the Kuder-Richardson Formula 20 was 0.77.

Among the validation formulas, Flesch values were inversely related to readability. This means that higher Flesch measures correspond to easier readability and lower Flesch measures correspond to more difficult readability. Similarly, high reading comprehension scores were expected for easier reading materials and low scores for difficult reading materials.

On the other hand, the Fry Readability Graph (Fry), the Communication Index (CI), the Judged Readability Level (JRL) and the JB formula yield measures which directly relate to readability. Higher readability measures correspond to difficult reading and lower readability measures correspond to easier reading.

These considerations led to the following expected results: Flesch measures and reading comprehension scores relate negatively with Fry, Communication Index, Judged Readability Level and the JB formula. Flesch and comprehension scores, however, relate positively with each other. The Fry, Communication Index, Judged Readability and the JB formula values, likewise, relate positively with each other.

If the results were consistent with these and were significant, the relationships are confirmed and would intervalidate the formulas. Significant relationship of any, or more, of these measures with the JB formula would, in turn, validate the latter.

RESULTS

As shown in Table 13, all the readability formulas, the judgment of readers, and students' test scores showed a complete agreement on the easiest and the most difficult passages.

Passage 1 was predicted as easiest and Passage 2 was predicted as most difficult. Flesch and Fry formulas had complete agreement in their readability level rankings of the five passages. Likewise, Feedback-Based Communication Index and the JB formula had complete agreement in their readability level rankings of the five passages.

Table 13
Predicted Readability Levels by Different Formulas and Average Scores of Students in Five Passages

	Predicted Readability Levels and Average Score					
Passage Number	FLESCH	FRY	CI	JRL	JB	Ave. SCORE
1	90	4	.00179	1.30	6.60	2.30
2	27	15	.01972	3.78	10.91	1.66
3	48	13	.00838	3.00	8.83	2.13
4	84	5	.00828	3.40	8.34	1.71
5	59	9	.01589	3.51	9.20	1.07
	Ranks of Predicted Readability Levels and Average Scores					
Passage Number	FLESCH	FRY	CI	JRL	JB	Ave. SCORE
1	1	1	1	1	1	1
2	5	5	5	5	5	4
3	4	4	3	2	3	2
4	2	2	2	3	2	3
5	3	3	4	4	4	5

Columns 1 to 3 of Table 14 shows the correlation coefficients of readability levels predicted by the four formulas. The rank-order correlation showed moderately high relationship between the Flesch and Fry formulas, on one hand, and the JB formula, on the other. The correlation index was 0.80 in each case. A perfect correlation was observed between the Communication Index and the JB formula.

The lower portion of Table 14 shows the correlation coefficients using the raw data. As in the rank order correlation, high correlations were observed between the Flesch & JB formulas and Communication index & the JB formula. The correlation between the Fry formula and the JB formula was not significant.

The rank order correlation coefficient between the Judged Readability Level and JB formula was 0.80. The Pearson product moment correlation was also high at 0.88 but not significant.

The rank order correlation coefficient between the students' test scores and the readability levels predicted by the JB formula was moderately high at 0.80. Moreover, the correlation of the readability level predicted by the JB formula for each passage and the score of each student in each passage was significant, though, low at -0.21.

Table 14
Correlation Matrix of Readability Levels Predicted by Different Formulas and Scores of Student in Five Passages

	Rank-Order Correlations					
	FLESCH (1)	FRY (2)	CI (3)	JRL (4)	SCORE (5)	JB (6)
FLESCH	1.00					
FRY	1.00	1.00				
CI	.80	.80	1.00			
JRL	.60	.60	.80	1.00		
SCORE	.40	.40	.80	.80	1.00	
JB	.80	.80	1.00	.80	.80	1.00

	Pearson Product Moment Correlations					
	FLESCH (1)	FRY (2)	CI (3)	JRL (4)	SCORE (5)	JB (6)
FLESCH	1.00					
FRY	-.99 **	1.00				
CI	-.87	.79	1.00			
JRL	-.66	.61	.85	1.00		
SCORE	-.10 *	-.05	-.28 **	-.28 **	1.00	
JB	-.91 *	.86	.95 *	.88	-.21 **	1.00

N (Score, x) = 715
N (Others) = 5

DISCUSSION

Factors of Readability

An analysis of the variables loading significantly in Factor 1 revealed that three variables, namely, MEQ, PMW, and PWAP, have mathematical characteristics. MEQ is the presence or absence of mathematical equation(s). PMW is the combination of mathematical words, relational words and words that compare physical quantities while PWAP is the density of these combinations.

These variables were associated with quantitative concepts that required higher order thinking processes (Acuña, 1987). Complex representations such as mathematical equations were usually included in a text when words were meager and inadequate to convey the exact relationship of quantities. On the other hand, these quantitative concepts might also influence the sentence structure especially the idea density and complexity of the passage.

Another significant variable in Factor 1 is the average sentence length or ASL. ASL has been defined as a measure of sentence structure and idea density (Chall, 1958). Its inclusion to Factor 1 characterized by quantitative ideas and relationships could be explained by the effect of syllable density in a given sentence length. Since a mathematical equation was defined as a word, it contributed to a large number of syllables because each symbol was counted as a syllable. Because of this, sentences with mathematical words or equations were associated with longer sentences.

Factor 1, therefore, can be named as *Quantitative Relationship and Sentence Structure Factor.*

Factor 2 is characterized by the type of vocabulary used, in this case, technical vocabulary. The variables that loaded well into this factor were: proportion of derived concept words (PDCW), proportion of words specific to physics (PWSP), average word length (AWL), and presence/absence of graphics elements (GRA).

PWSP and PDCW were overlapping measures because they both required previous exposure or knowledge of physics concepts before reasonable understanding can take place. Thus, PDCW and PWSP can be said to carry not only of relatively complex ideas but also abstract ones, which in turn, contributed to the vocabulary difficulty as well as semantic difficulty and complexity of reading passages.

Normally too, technical terms required graphic illustrations to clarify. Therefore, the presence of graphics elements is an indication of the complexity of idea(s) carried by the vocabulary used in the textual passages. This is evident by GRA's inverse relationship with the factor.

Taking all these into consideration, Factor 2 could be named as *Technical Vocabulary.*

Factor 3 included proportion of personal, events, and places words and sentences (PPEPS and PPEPW). The distinctive characteristic of this factor was the inclusion of personal words, events, and places. These words were not usually included in content area texts like Science. They are, however, words which readers easily identify with and which could enhance interest. This factor may be named as *Affective Score*, similar to Flesch's (1948, 1951 as cited by Chall) Human Interest Score. The difference lay in the addition of events and places to personal words, word groups which were sparingly used in Physics and other similar content area books.

The above findings supported the multi-factor theory of readability. More specifically, that the readability of physics instructional materials is multi-factor. These factors are: *Quantitative Relationship and Sentence Structure*, *Technical Vocabulary*, and *Affective Score.*

Choosing the best readability formula was not easy. Ultimately, the basis for the choice did not entirely rest on the statistical formalism. The plausibility and perceived sensitivity of the formula to measure readability factors inherent in content area text like physics were considered.

While readability researchers generally recommend the sufficiency of two predictor variables in a readability equation, the present work opted for the inclusion of three. These variables were measures of sentence length and vocabulary (mathematical words and words specific to physics). The reason for this was not only due to the variables' significant contribution to prediction. It was also based on the pretext that concept formation for abstract technical concepts, like those that were embraced by the two vocabulary variables in the equation, is long and tedious. Thus, including three variables can help focus on the special features of science instructional materials which are characterized by the incidence of technical vocabulary.

The best single predictor variable, average sentence length, (ASL) commonly appears in numerous readability formulas. This variable invariably relates positively with readability. This is because long sentences usually carry more ideas or words and are likely to be more complex in their sentence structure.

Average word length (AWL) which prominently figured in other readability formulas like the Flesch and Fry did not come out significant in the present work. Instead, other variables, proportion of mathematical words (PMW), and proportion of words specific to physics (PWSP), came out as better predictors of readability. These variables were found in other formulas. The entry of these variables in the present formula could make it a legitimate measure of readability for a content area like Physics. Being a specialized field, Physics is fraught with distinctive words like those that entered the formula. Both groups of words require previous exposure for an accurate understanding.

Mathematical words have a wide range of idea complexity, from a simple and common concept like addition, circle or line to more sophisticated and complex process words such as integral or exponential. Words specific to physics (PWSP) carry technical meanings. These are usually abstract and subtle and require some time for concept formation.

This can also be true to other content areas that require extensive use of mathematics as a medium for developing its internal concepts. Physics as a special subject area progressively introduces mathematics along with the development of ideas. Physics is rigorously associated with mathematics which is extensively used as a tool in developing physics ideas and concepts. Most Physics concepts are, in fact, mathematical in nature.

Validity of the JB Readability Formula

Among the formulas, the Flesch and Fry as a pair measured readability very similarly. The results also indicate that the Communication Index formula, the JB formula as well as the Flesch formula measured the same thing. Between the Flesch formula and the JB formula, the significant relationship can be attributed to a common variable, average sentence length (ASL).

The significant relationship between the Communication Index and the JB formula could be indicative of the latter's sensitivity to identify words and sentence factors which agree with what readers consider as unclear. These are technical vocabulary represented by PWSP, quantitative relations represented by PMW and sentence factor represented by ASL.

Previous results by Talisayon (1983) showed a different result where the Communication Index had little or practically no correlation with the Flesch and Fry formulas. The study explained that factors like reader characteristics gleaned through reader feedback could play a significant role in estimating readability beyond sentences and words, which can be especially true for content materials like Physics.

A deeper insight is gleaned by taking a closer look at the unclear elements identified by students during the process of determining the Feedback-based Communication Index of the passages. A considerable portion of the words marked unclear by the students were not words, per se, but a combination of words which, when taken together, formed a specific technical concept. Examples of this include "altitude profile", "hydrostatic balance", "radiation inversion", "centrifugal force", "radius vector" and "potential energy". It can be noticed that many of the words in such phrases, when taken independently, carry commonly accepted layman meanings, but takes on a specific technical meaning when combined with the other word.

Many of the phrases marked as unclear were either mathematical words or words specific to physics which represent quantitative relationships and technical vocabulary. Similarly, the sentences which were marked unclear by the students were characterized by their length, such as:

> *"In hydrostatic balance, the atmospheric pressure at any height equals the total weight of overlying gas, a condition which requires that the pressure and density of the gas decreases exponentially at a rate inversely proportional to the temperature."*

This sentence was marked by five students, the highest frequency among the unclear sentences marked. The sentence is, likewise, characterized by the presence of mathematical words and words specific to physics.

The fact that students marked whole sentence and, at times, whole paragraph as unclear indicates that readability goes beyond sentence length, vocabulary, and idea relationships. This may be explicit in the JB formula. The high and significant relationship between the JB formula and the Feedback-based Communication Index could, however, indicate that the quantitative predictor variables in the present formula may be capable of encompassing implicit factors beyond their explicit definitions. If this is so, then these variables combine to obtain excellent approximations of the true measure of readability.

The significant relationship of the JB formula with scores in the reading comprehension test can be explained by the common notion that easier reading materials are easier to understand which, in turn, translates to higher reading comprehension. Difficult reading materials, on the other hand, are harder to understand and impedes reading comprehension.

The above significant relationships show the ability of the formula to classify passages according to difficulty. These results also demonstrate the capability of the developed formula (JB formula) as a valid model for measuring readability.

Application of the JB formula, however, also entails some limitations as with the other readability formulas. The formula intercept has a magnitude of 5.5516 which limits the formula's sensitivity to readability levels below this mark. It was also formulated based on the average reader's perceived readability level, hence, the formula may not necessarily appeal to the specialist's or expert's demands. In addition, the formula does not include measures of readability variables as writing style, organization, syntax and the like.

Therefore, the formula can only calculate readability estimates not absolute readability levels.

Procedure in Using the JB Formula

1 Randomly select passages of at least 100 words. The passages should preferably be a complete paragraph or paragraphs depicting a complete idea.

   To determine the readability of an entire book, take at least one or two passages from each chapter. To determine the readability of a chapter, take at least one passage for every 10 pages.

2. Count the number of syllables in the passage (refer to the Theoretical Framework section for the operational definition of a syllable).

3. Count the number of sentences in the passage. Determine the average sentence length (ASL) using:

$$ASL = (\text{no. of syllables/no. of sentences})$$

4. Count the number of words.

5. Count the number of words specific to physics (WSP). Solve percentage of words specific to physics (PWSP) using:

$$PWSP = 100(\text{no. of WSP})/(\text{no. of words})$$

6. Count the number of mathematical words. Then compute the percentage of mathematical words (PMW) using:

$$PMW = 100(\text{no. MW})/(\text{no. of words})$$

7. Compute the Reading Difficulty Score using:

$$RDS = 5.5516 + 0.0949\ ASL + 0.0951\ PMW + 0.0506\ PWSP$$

8. Classify the passages/chapter/book using the following scale:

Grade Level	Readability Level
Elementary Grades	Less than 6.50
First & Second Year, High School	6.51 - 8.50
Third & Fourth Year, High School	8.51 - 10.50
First & Second Year, College	10.51 - 12.50
Third Year, College and Beyond	Beyond 12.50

REFERENCES

Acuña, J. E. and De Guzman, F. S. (1987). Language and science achievement in Philippine context. *Journal of Science and Mathematics Education in Southeast Asia. 1*, 20-27.

Cañares, A. (1991). *Readability of selected introductory physics textbooks*. Unpublished master's thesis, De la Salle University.

Chall, J. S. (1958). *Readability: An appraisal of research and application.* Columbus, Ohio: The Ohio State University.

De la Cruz, V. (1966). *A comparative study of the readability of elementary science textbooks*. Unpublished master's thesis, University of the Philippines.

Espartero, A. (1976). *A comparative study of the readability of secondary science textbooks*. Unpublished master's thesis, University of the Philippines.

Fry, E. B. (1968). A readability formula that saves time. *Journal of Reading. 11*(7), 513-516.

Fry, E. B. (1977). Fry's readability graph: Clarification, validity, and extension to level 17. *Journal of Reading. 21*, 242-252.

Galitano, L. (1974). *Readability of some elementary mathematics textbooks*. Unpublished master's thesis, University of the Philippines.

Grouws, D. & Robinson, R. (1973). Some ideas concerning the readability of classroom mathematics materials. *School Science and Mathematics*. 7(9), 711-716.

Hair, J. F., Anderson, R. E., & Tatham, R. L. (1990). *Multivariate data analysis with readings (2nd ed.)*. New York: McMillan Publication Co..

Lagarde, C. (1984). *A comparative study of the readability of selected secondary physics textbooks*. Unpublished manuscript, Philippine Normal College.

McConnel, C. R. (1982). Readability formula as applied to college economics textbooks. *Journal of Reading*. 14-17.

Norusis, M. J. (1988a). *SPSSPC+ Advanced Statistics*. Version 2.0 for IBM PC/XT/AT and PS/2.

Norusis, M. J. (1988b). *SPSSPC+ Base Manual*. Version 2.0 for IBM PC/XT/AT and PS/2.

Perez, C. (1982). *Readability level determination for elementary science texts* (Monograph No. 27). University of the Philippines, Science Education Center.

Spiegel, D. L. & Wright, J. D. (1984). Biology teacher's preferences in textbook characteristics. *Journal of Reading*. 624-628.

Talisayon, V. M. (1983). *Feedback-based readability formula for science and mathematics curriculum materials* (Monograph series). University of the Philippines, Science Education Center.

Young, F. (1991). *Development and evaluation of modules on selected topics in statistics*. Unpublished master's thesis, De la Salle University.

LANGUAGE AND UNDERSTANDING IN MATHEMATICAL PROBLEM SOLVING

Allan B. I. Bernardo
Department of Psychology
University of the Philippines

Abstract

This research is an investigation on the effects of language factors on two specific components of understanding in math word problem solving: the comprehension of the text information and the construction of the problem structure. In two experiments, Filipino-English Bilingual students were given arithmetic word problems in either Filipino or English. In Experiment 1, a recall procedure was used to assess the students' comprehension of the problem texts. The results show that students comprehend the problems better when it was in their first language; students were also more accurate in solving the problems in their first language. The results suggest that language proficiency facilitates the first stage of understanding word problems and also the overall problem-solving process. In Experiment 2, a problem-completion procedure was used to assess the students' ability to construct the problem structure. The results show that the students constructed the problem structure equally well in either language. The results suggest that language factors do not affect the more abstract stage of processing word problems. The dissociation of language effects on the two components are discussed in terms of the theoretical distinction between levels of understanding in word problem solving. The implications of the results on the assessment of the relationship between language use and mathematical performance, particularly between language use and mathematical understanding and learning among bilinguals were also discussed.

What is the relationship between language skills and mathematical abilities? Researchers in educational achievement and psychological measurement in the Philippines have long known about the robust correlation between performance in English language classes and in mathematics classes. On the other hand, there seems to be a weak correlation between performance in Filipino classes and in mathematics classes (Department of Education and Culture, 1976). These findings seem to suggest some link between English language skills and mathematical abilities, but not between Filipino language skills and mathematical abilities. One could speculate that this link is related to the English language's greater efficacy in expressing mathematical concepts and operations. It is possible that the structures of the English language affords better handling of the abstract information in mathematics. (It was probably this view that led the proponents of our country's Bilingual Education Policy to require that English be used as the medium of instruction for teaching mathematics in particular).

However, such interpretation of the correlation data should be accepted with great caution; in fact, alternative interpretations are just as plausible (e.g., tests of English verbal ability and of mathematical ability require the same type of test-taking competencies, while tests of Filipino require different types).

There is also a general constraint regarding interpreting results based on psychometric tests of human abilities. This constraint regards the specific psychometric definition of the various abilities that are being studied. The constraint is that tests of mathematical proficiency and ability, for example, do not provide an independent description of what is being measured (Mayer,1985). Hence, it is difficult to make specific claims about particular relationships between abilities in various domains unless there is a clear delineation of the nature of the knowledge and skills that underlie the various abilities in the first place.

The purpose of this research was to investigate the relationship between language and mathematical ability using an information-processing approach to studying performance in simple word problems in arithmetic. The information-processing approach specifies information-processing components involved in the task of solving such word problem. The research investigated the particular function that language may play in the operation of these components. By using an approach that characterizes mathematical ability in terms of specific component abilities, the research aims to develop a more detailed picture of the relationship between language and mathematical performance.

The Information-Processing Approach

The information-processing approach to studying various human abilities is primarily based on a detailed analysis of the tasks of the different domains (Sternberg, 1977, 1985). The assumption of the approach is that any type of task or problem in any domain can be broken down into information-processing components. These components are the various simple mental processes or operations, skills, and knowledge that are required for executing the tasks or solving the problems in a domain.

Mayer (1985) described the general form of an information-processing analysis of mathematical performance, and this basic approach has been utilized in entire programs of research on mathematical performance (see e.g., Schoenfeld, 1985). According to Mayer, mathematical problem solving can be broken down into two major components: *problem representation* and *problem solution*. Problem representation refers to the process of deriving a mental representation of the problem elements from the problem text, while problem solution refers to the process of applying different mathematical operations to the mental representation in order to arrive at a final answer. Both components are equally important in determining performance; in fact, error in either component can lead to poor problem-solving performance. The use of inappropriate math operations or the incorrect use of appropriate operations on the mental representations will lead to the wrong problem solution. Similarly, beginning with an incorrect mental representation of the problem will lead to the wrong solution, even if the correct math operations are executed flawlessly.

An Information-Processing Model of Math Word Problem Solving

Models of arithmetic word problem solving using the information-processing approach have been proposed by Briars and Larkin (1984), by Riley, Greeno, and Heller (1983), and others. These models assume that the problem-representation component of solving word problems involves the development of structural problem representations or problem-type schemata and using these schemata to guide the comprehension of the abstract problem elements. The problem solution would then be based on the understanding of the structural elements of the problem.

The model of word problem solving that was used in the study adds at least one important element to the models just described. Kintsch and Greeno's (1985) model assumes that aside from the structural problem-solving aspects, there is also a text-comprehension component in word problem solving. Kintsch and Greeno proposed that the task of understanding the text of a math word problem involves the construction of a conceptual representation of the structural elements of the problem from the verbal form of the problem. In particular, they proposed a dual representation which includes a *propositional text base* and a *problem model*. The propositional text base represents the information in the text. The problem solver transforms the verbal input into a list of conceptual propositions representing its meaning (e.g., people and objects mentioned, how they are related to each other, the overall story, etc.) These propositions are organized such that the general concepts (e.g., sets and set relations) are made salient. The problem model, on the other hand, refers to the structural information needed to solve the problem. In constructing the problem model, the problem solver begins with the information in the text base. The problem solver sorts out this information, excludes information that is not required for the solution, and infers information that is needed for solving the problem. After the problem model is constructed, the problem solver then uses a set of counting and arithmetic operations for calculating the solutions of the problem.

An important feature of the Kintsch and Greeno model is the primacy given to text-comprehension processes, which is fundamentally a linguistic activity. Hence the model allows for the effects of what are basically linguistic variables on math problem-solving performance. That linguistic variables might have an effect is particularly significant considering that mathematical activity is often thought to be an abstract procedure that is not supposed to be affected by non-structural or non-abstract factors.

Language Factors in Math Problem Solving: The Empirical Evidence

Consistent with the Kintsch and Greeno model, several researches have shown that variations in the given text of the problem leads to systematic effects in problem-solving performance, even if as such variations do not affect the underlying problem structure. Several studies showed that changing the linguistic structure of word problems lead to substantial differences in successful problem solution (see e.g., Davis-Dorsey, Ross, & Morrison, 1991; DeCorte & Verschaffel, 1987; DeCorte, Verschaffel, & Pauwels, 1990; Nesher & Katriel, 1977; Nesher & Teubal, 1977). For example, DeCorte, Verschaffel, and De Win (1985) found that children performed better on problems in which the wording was made more explicit in terms of set and set relations compared to the typical sparse wording in arithmetic problems. Research by Cummins, Kintsch, Reusser, and Weimer (1988), and by Riley and Greeno (1988; Riley, Greeno & Heller, 1983) showed that difficulties in comprehending and remembering the texts of word problems lead to more errors in problem solutions. The findings of all these studies suggest that the comprehension or linguistic understanding of the problem texts affects the process of developing a mental representation of the problem information, and hence, also affects problem solving performance.

Research on analogical transfer in math problem solving also suggests a possible role of language factors. For example, research by Reed, Dempster, and Ettinger (1985) on word problem solving in algebra showed that students primarily relied on superficial similarities between problems in transferring information from one problem to another. In other words, instead of using structural similarities between problem as the basis for transferring information, students transfer problem information based on superficial features of the problem like similarities in the story (see also, Bernardo & Okagaki, 1994; Novick & Holyoak, 1991; Ross, 1984, 1989). In a study of word problem solving in probability among bilingual subjects, Bernardo (1994a) even showed that such transfer of information is more likely when the analogical problems are written in the same language than when they are written in different languages.

A Model of Language Use and Understanding in Math Problem Solving

Among the various aspects of word problem solving, the focus of the study was on the component of understanding word problems because problem understanding is equivalent to the process of forming a problem representation. One of the most basic principles of psychology of problem solving is that the correct problem solution is dependent on the formation of the correct problem representation (see e.g., Duncker, 1945; VanLehn, 1989). It is important to note that most of the studies reviewed in the previous section showed related effects to the process of problem representation.

To reiterate the assumption of Kintsch & Greeno's theory, problem understanding involves two components: the construction of a propositional text base and the construction of an abstract problem model. The first is a more basic, although just as important, mode of understanding which refers to whether or not a student has an accurate comprehension of the situation that is being described in the word problem. The second component of understanding refers to whether or not the student grasps the quantitative relations stated in the problem, the pertinent mathematical principles that underlie these relationships and so on. Both forms of understanding are necessary for proficient performance in word problem solving.

The general thesis of this research is that language has a specific function that relates to each of these components of understanding in word problem solving. Therefore, the research hypothesis is that there is no unitary effect of using the first or the second language in mathematical understanding. The effect of language factors will be different for each component of understanding.

In particular, it is hypothesized that using the student's first language will result to better problem text comprehension than using the student's second language. This hypothesis is based on the notion that the construction of a propositional text base is dependent on the individual's capacity to correctly parse sentences in a language. Therefore, prose comprehension should be better in the individual's first language or more proficient language.

On the other hand, the specific language used in the problem should not have any effect on the structural understanding of the problem, regardless of whether the language is the student's first or second language. This hypothesis is based on the view that the knowledge used for constructing the problem structure is fundamentally abstract, and hence, should not be dependent on specific language representations.

Overview of Experiments

Two experiments were designed in this study to investigate the effects of language on the two components of understanding in word problem solving. Experiment 1 assessed the effects of language factors on the first level of understanding math word problems -- text comprehension; Experiment 2 assessed the effects of language factors on the second level of understanding -- problem-structure construction.

In Experiment 1, Grade 2 students were presented word problems in arithmetic in Filipino and in English. The students have been just introduced to solving math word problems, and have not had extensive experience with such problems (therefore, they have not developed elaborate memory representations for such problems, c.f., Bernardo, 1994b; Ross & Kennedy, 1990). They were asked to recall each word problem after it is presented to them. The recall task should reveal how well the student has comprehended the material in the text. The assumption is that accuracy of recall as well as the patterns of errors in recall reflect a student's textual understanding. The experiment involved students from public schools in Quezon City. These student have Filipino as their first language, and they also have relatively poor proficiency in English. However, these students are being taught mathematics using the English language and using materials written in English. All students are taught mathematics using the English language and using materials written in English.

In Experiment 2, Grade 4 students were also presented math word problems in English and Filipino. However, these problems were incomplete or the question at the end of each problem was omitted. The students were required to complete the problem. These students have had extensive experience with such problems, and therefore, should have acquired some level of proficiency in constructing the problem structure for these problems. It was assumed that the students' ability to complete a problem would primarily depend on the students' ability to correctly construct the problem structure. The experiment involved students from a public school in Quezon City, similar to those in Experiment 1.

In summary, Experiment 1 used a recall paradigm to test the hypothesis that the text-comprehension component of word problem solving will be better in the students' first or more proficient language. Experiment 2 used a problem-completion paradigm to test the hypothesis that the problem-model construction component of word problem solving is not affected by language factors.

Experiment 1

This experiment was designed to investigate the effects of the language used in math problems on the specific problem component of text comprehension among bilinguals. To assess whether the subjects comprehended the story information in word problem, subjects were required first to recall each problem, and then solve the problem they recalled. This same procedure was used by Cummins et al. (1988) to demonstrate textual understanding of word problems. The assumption is that the subject's recall would be based on the memory representations formed after encoding the text information. Furthermore, better memory representations would be formed if the subject has a more complete and accurate understanding of the text information.

The subjects, who were Filipino-English bilinguals, were given problems written either in Filipino or in English. Filipino was the subjects' first language, and English was their second language. However, all these subjects are having their mathematics education using the English language and materials written in the English language. By comparing the subjects' performance on the Filipino problems to their performance on the English problems, we can determine the effects of the language proficiency on the first important component of word problem solving. There are possible effects that could have been observed. The subjects could have performed better in their first language, Filipino. They could have performed better in the language in which they acquired the mathematical knowledge, English. Or they could perform equally well in both languages. However, it was hypothesized that subjects would have better recall for problems in Filipino, their first language. This hypothesis was based on the notion that since comprehension of text is a basic linguistic skill, subjects should be more efficient when processing linguistic information in their first language or the language in which they are most proficient. It was further hypothesized that the pattern of recall performance would parallel solution performance. That is, subjects would also provide more correct solutions for problems in Filipino. This hypothesis is based on the model of word problem solving discussed earlier in which text comprehension is an integral part of the complete word problem solving process. Therefore, performance in that component should have a corresponding effect on the complete task.

Method

Subjects. The participants in this experiment were 43 Grade 2 students from the two public schools in Quezon City. All the students reported that they used Filipino at home, but can speak, understand, and read enough of English as it is used in their school work. The students were taught mathematics in English, and were just introduced to arithmetic word problems.

Materials. The problem used in this experiment were based on the 18 story problems used by Riley, et al. (1983). These problems are presented in the Appendix. They consist of six specific problems within each of the three major problem types. The problem types are as follows: Combine problems, in which a subset or superset must be computed given information about two other sets; Change problems, in which a starting set is changed by transferring items in or out, and the cardinality of the starting set, transfer set, or results set must be computed given information about two of the sets; Compare problems, in which the cardinality of one set must be computed by comparing the information given about the sets. All the problems included only the numbers 1 through 9, and the correct answers ranged only from 1 to 10. The word problem used in this experiment all contained "Judy and Carlo" as actors and "candies" as objects. This procedure was done to reduce the memory load requirements of the task; that is, the students only had to attend to the numbers stated in the problem and their described relationships.

The 18 word problems were first written in English and then translated into Filipino by a research assistant who was fluent in both Filipino and English. The Filipino translations were translated back into English by another research assistant to ensure that the translations were equivalent. Each student worked on 18 problems, 9 in Filipino and 9 in English. For this purpose, the 18 problems were divided into two sets (Sets A and B) with each having 3 word problems from each of the three problem types. For about half of the subjects Set A was given in Filipino and set B in English, for the other subjects Set A was given in English and Set B in Filipino. Three different random arrangements of each set were created and the assignment of each random sequence was counterbalanced across subjects. Likewise, the assignment of which language set to work on first was counterbalanced across subjects.

Based on earlier research using these 18 word problems (e.g., Cummins, et al., 1988), 4 problems were identified as easy problems and the other 12 as difficult problems. The 4 easy problems are the first two problems in both the Combine and Change problem types. The assignment of easy and difficult problems from the three problem types to Sets A and B were balanced. Therefore, each set of 9 had 1 easy and 2 difficult Combine problems, 1 easy and 2 difficult Change problems, and 3 difficult Compare problems.

In addition to the Filipino and English versions of the word problems, the numeric format of the problems were also used. For example, for the fifth Change problem in the Appendix, the corresponding numeric format was, $\underline{?} + 3 = 5$. Each numeric format problem was presented in the horizontal form described in the earlier example. The 18 numeric format problems were presented to subjects in one random sequence on a sheet of paper.

Design. Two variables were manipulated in this experiment: the language of the problem (Filipino or English) and the difficulty of the problem (easy or difficult). The two factors were studied using a 2 x 2 completely repeated factorial design. There were two dependent variables measured: the proportion of correct recall and the proportion of correct solutions for the problems for each of the factorial conditions.

Procedures. The student were tested individually in a reasonably quiet place in their school during school hours. All the problems were presented orally; each problem was read to the student twice, or thrice if the student requests. Whenever possible, the student's responses were tape recorded. All students were told that only the researcher would hear the tapes, and that their teachers and parents would not. No student objected to her sessions being taped, although in some cases tape recording was not possible because of technical problems.

Each session began by getting acquainted with the student and asking the student which language she preferred to use in the session. She was then asked to write her name and birthday on a sheet of paper; in case the student could not recall the year of her birth, she was asked how old she was on her last birthday. The student was then asked what language she used at home. She was informed that they would be playing some kind of a game involving the word problems. Examples of word problems were then presented and she was asked to try recalling them. She was told that they would be doing more of the same procedure and that for each problem she will listen to and recall, she will also be asked to compute for the solution on the sheet of paper with her name. After the student completed the Filipino and English sets of 9 problems each, she was given the sheet of paper with the 18 numeric format problems. All students were given gifts as tokens for their participation.

Results and Discussion

Recall Data. Each of the responses for all of the problems was coded as showing correct comprehension or not. Correct comprehension was scored if the subject recalled the problem verbatim or the subject paraphrased or transformed the problem without altering the object relations described in the text. The mean proportion of correct recall for Filipino and English, easy and difficult problems are summarized in Table 1.

Table 1
Mean proportion of Correct Recall (and Standard Error) as a Function of Problem Difficulty, and Language of Problem in Experiment 1

	Language of Problem			
Difficulty of Problem	Filipino		English	
Easy	80.2%	(4.7)	51.2%	(5.4)
Difficult	20.6%	(5.4)	11.0%	(2.3)

The means were analyzed using a 2 x 2 Analysis of Variance (ANOVA) for completely repeated factorial design. The analysis showed a main effect of the language of the problem, $F(1, 42) = 28.20$, $MSe = .057$, $p<.0001$, suggesting that subjects correctly understood the problem texts in Filipino more often than the problems in English. There was also a main effect of problem difficulty, $F(1, 42) = 141.07$, $MSe = .076$, $p<.0001$; subjects recalled more of the easy problems compared to the difficult problems. There was also a significant interaction between language and difficult of the problem, $F(1, 42) = 7.84$, $MSe = .052$, $p<.008$; the advantage for Filipino problems was more marked in the easy problems than in the difficult problems. However, as Table 1 shows, this interaction might be due to the rather low recall rates for the difficult problems which might have led to some floor effect.

As hypothesized, subjects performed better on the Filipino problems compared to the English problems. Subjects understood the problem texts when it was written in the language in which they are most proficient. This was true even if the subjects are being instructed in math using the English language.

Solution Data. The subjects' computed solutions were also coded for accuracy. The mean proportions of correct solutions are summarized in Table 2. A notable result was that when all the easy and difficult word problems were rendered in numeric format, subjects performed very well-- 97.7% accuracy for easy problems and 95.2% for difficult problems. These accuracy scores are notable higher than the accuracy scores for the word problems either in Filipino or in English, a result that replicates Cummins, et al. (1988). These results underscore the effect of linguistic processing in math word problem solving.

Table 2
Mean Proportion of Correct Solutions (and Standard Error) as a Function of Problem Difficulty and Language of Problem in Experiment 1

	Language of Problem			
Difficulty of Problem	Filipino		English	
Easy	87.2%	(3.4)	73.3%	(5.1)
Difficult	41.2%	(3.1)	36.9%	(3.3)

The means were also analyzed using a 2 x 2 ANOVA for completely repeated designs. Similar to the recall data, there were also main effects of language of test, $F(1, 42) = 8.10$; $MSe = .044$, $p<.007$, and of difficulty, $F(1, 42) = 103.60$, $MSe = .070$, $p<.0001$. Subjects arrived at the correct solutions more often when the problem was in Filipino, and when the problem was easy. However, the interaction effect was not statistically reliable, $F(1, 42) = 1.80$, $MSe = .056$, $p<.10$. This suggests that the effect of difficulty on subjects' solutions was the same whether the problems were in Filipino or in English. The most significant finding in these results, however, is that the advantage that the students had in understanding the texts of problems in Filipino corresponds to a similar advantage the subjects had in solving problems in Filipino. These findings provide support for the view that since text comprehension is an important part of word problem solving, success in text comprehension should lead to success in solving the word problems. More importantly, subjects understood and solved the problems better if the problems were written in their first language or the language in which they are most proficient.

Qualitative Analysis: Type of Recall. To further explore the relationship between the language of the problems, recall performance, and solution performance, the scores were recoded and analyzed quantitatively. Each recall response was recoded into one of the following categories: correct verbatim recall (VR), structure-preserving transformations (SP), structure-violating transformations (SV), nonsense problem (NP), or others (OT). The frequency of responses in each category (across subjects and problem difficulty) for Filipino and English problems are summarized in Table 3.

Correct verbatim recall (VR) referred to responses that when practically verbatim constructions of the problem text -- there were at most, two minor differences in the wording. This type of response accounted for 25.6% of the responses to the Filipino problems and 14.5% of the responses to the English problems. These results already suggest an advantage of using the student's first language for recalling and understanding the problem text.

Responses were coded as structure-preserving transformations (SP) when the wording of the problem was significantly changed during recall,

Table 3
Frequency of Recall Responses as a Function of Language of Problem in Experiment 1

Recall Type	Language of Problem	
	Filipino	English
Correct Verbatim (CV)	99	56
Structure Preserving (SP)	34	29
Structure Violating (SV)	80	85
Nonsense Problem (NP)	162	200
Other (OT)	12	17

Note: Frequencies are based on 9 observations for each set of Filipino and English problems for each of 43 subjects.

but the important quantitative relations between sets was maintained. An example of an SP is when the following Compare problem:

> *Judy has 9 candies. She has 4 candies more than Carlo. How many candies does Carlo have?*

was recalled as:

> *Judy has 9 candies. Carlo has 4 candies less than Judy. How many candies does Carlo have?*

(SP and VR were coded as correct recall in the original recall coding because both responses indicated correct understanding of the problem text.) There were only a few such transformations; 8.8% and 7.5% of the responses for the Filipino and the English problems, respectively were coded as SP's. The fact that there seems to be no difference in SP's for Filipino and English problems, suggests that the advantage in understanding Filipino problems was due to an advantage in the direct parsing of the problem texts in Filipino compared to English (as shown by the VR data), rather than to an elaboration of the problem text information (which would have been revealed by the SP data).

Responses which included significant wording changes that alter the mathematical relationships were coded as structure-violating transformations (SV). An example of this is when the compare problem mentioned above is recalled as:

> *Judy has 9 candies? Carlo has 4 candies? How many do they have together?*

These responses accounted for 20.7% of the Filipino problems and 22.0% of the English problems. Again, there seems to be no difference in the tendency to make this error in either language. That is, as far as misunderstanding the text information in the problem, there seems to be an equal likelihood for English and Filipino problems.

The next category of responses is the nonsense problem (NP). The problems are recalled problems that simply ask for the numbers given in the problem. The following are two examples:

Judy has 7 candies. Carlo has some candies. How many candies does Judy have?

Judy and Carlo have 8 candies altogether. Judy has 7 candies. Carlo has some candies. How many candies do they have together?

This error indicates a basic flaw in the understanding of the problem, in the sense that the recalled problem does not even reflect the most fundamental relationships among the quantities in the text. In other words, the error reveals that the respondent had a very superficial understanding of the information in the text. A total of 41.8% of the responses for the Filipino problems and 51.7% of those for the English problems were NP's. There seem to be more of this type of severe error when subjects are solving English problems.

Finally, responses that did not fit any of the earlier categories were coded into a catch-all other (OT). The OT's included recall of problems that were required non-qualitative solutions (e.g., "How many candies can Judy give her brother?" or "Why does Judy have more than Carlo?"), partial recall (e.g., "Judy and Carlo have 8 candies altogether. Judy has 7 candies..."), no answers, and other types of responses. a total of 3.1% of the responses for the Filipino problems and 4.4% for the English problems were coded in this category.

What this preliminary qualitative analysis shows is that the advantage in recalling Filipino problems seems to be occurring at the level of very basic comprehension skills. That is, the students are better at the fundamental parsing of text in Filipino compared to English (c.f., the VR's) and more likely to loose track of the over-all sense of the text in English compared to Filipino (c.f., the NP's). That the effect of language seems to be at a most basic level reflects the lopsided proficiency of the students in Filipino compared to English. The significance of this qualitative finding is underscored if we consider the solutions that students give for each type of recall response.

Qualitative Analysis: Recall and Solution Performance. For further qualitative analysis, the solutions that the subjects gave for the problems were also recoded using a more specific scheme. Each solution was coded as either a correct solution (CO), a wrong operation error (WO), a given number error (GN), and arithmetic error (AE), or an unclassifiable error (OT). To illustrate, consider the following problem:

Judy and Carlo have 8 candies altogether. Judy has 7 candies. How many candies does Carlo have?

A CO solution was "8 - 7 = 1". A WO error was "8 + 7 = 15" (operation used was addition instead of subtraction), while a GN error "8 - 6 = 2" (the number subtracted was 6 instead of 7). A solution of "8 - 7 = 3" was an AE, while a solution of "3 + 4 = 9" was an OT. The WO and GN errors are important errors for the analysis, because they reflect errors in understanding the problem text. On the other hand, AE or OT reflected problems with knowledge about basic numerical informations. The frequency of each type of solution corresponding to each type of recall response is summarized in Table 4.

Table 4
Frequency of Recall and Solution Responses in Experiment 1

	Solution Type					
Recall Type	CO	WO	GN	AE	OT	TOTAL
FILIPINO PROBLEMS						
Correct Verbatim (CV)	83	7	2	1	6	99
Structure Preserving (SP)	18	8	4	0	4	34
Structure Violating (SV)	26	34	1	0	19	80
Nonsense Problem (NP)	67	46	9	0	40	162
Other (OT)	6	2	1	0	3	12
Total	200	97	17	1	72	387
ENGLISH PROBLEMS						
Correct Verbatim (CV)	48	5	0	1	2	56
Structure Preserving (SP)	18	9	1	1	0	29
Structure Violating (SV)	30	31	7	0	17	85
Nonsense Problem (NP)	81	55	11	1	52	200
Other (OT)	6	2	0	0	9	17
Total	183	102	19	3	80	387

Note : Frequencies are based on 9 observations for each set of Filipino and English problems for each of 43 subjects. CO= correct solution; WO= wrong operation error; GN= given number error; AE= arithmetic error; and OT= unclassified error.

The data show that when the subjects correctly recall the problem verbatim, they are likely to be also correct in solving the problem. Of the 99 correct responses for the Filipino problems, 83.8% led to correct solutions; and of the 56 correct recall responses for the English problems, 85.7% led to correct solutions. Since there were more instances of correct recall for the Filipino problems, overall, there was better solution performance for these problems.

If one looks at nonsense problems, a coherent picture emerges. Of the 162 nonsense problems in Filipino, only 41.4% led to correct solutions, while 34.0% led to errors that could be associated with miscomprehensions (WO's & GN's), and 24.7% led to errors associated with knowledge about number operations (AE's & OT's). Similarly, of the 200 nonsense problems in English, only 40.5% led to correct solutions, 33.0% led to errors that could be associated with miscomprehensions, and 26.5% led to errors associated with knowledge about number operations. Since there were more nonsense problems in English, overall, there were more errors in the English problems, too.

This particular qualitative analysis suggests that the language effects at the basic level of comprehension that were found in the earlier quantitative analysis are closely related to the solution performance of the subjects. It seems that the overall better solution performance with the Filipino problems can be traced to the more effective basic parsing of Filipino problems and greater tendency to get lost in the parsing of English problems.

To summarize the results, the data of Experiment 1 provide evidence for the hypothesized relationship between language and a specific component of math word problem solving -- text comprehension. The results of the parametric analyses showed that understanding of the problem text was better for the problems written in the student's first or more proficient language. Furthermore, solution performance was also better for the problems in the student's first language. The results of the qualitative analyses showed that these language effects seem to be because of very basic level differences in processing proficiency between the two language. Furthermore, these difficulties associated with understanding problems lead to corresponding errors in solutions.

Overall, the results of Experiment 1 suggest that for the process of understanding the textual information in the word problems, there needs to be a match between the most proficient language the problem solver has and the language of the problems that need to be solved. The results specify a particular function of language; that is, language skill as a tool for understanding materials in word problems in math. However, this specific role of language in mathematical performance is predicted to be specific to the component of text comprehension. The role of language might be altogether different in other components of math word problem solving as will be seen in Experiments 2 that look into the component of constructing the problem structure.

Experiment 2

In this experiment, we studied a different component of word problem solving: the construction of the problem structure. According to theories of word problem solving (Kintsch & Greeno, 1985), after the problem solver comes to understand the textual information and constructs the text base, she then begins to construct the abstract problem model from the text base. The problem solver uses knowledge about typical problem solving operations to construct an abstract structural representation of the problem.

The subject's ability to construct the problem structure was assessed by presenting subjects with incomplete problems. The same problems used in the earlier experiments were presented without the question. The subjects were asked first to complete the problem by providing the correct question and then to solve the problem as completed. The same procedure was used in earlier studies by Krutetskii (1976) and Cummins, et al. (1988). The assumption is that for these problems, problem solvers can logically infer the question from the given problem information if they grasp the quantitative relationships that underlie the problem structure.

Similar to Experiment 1, the subjects were Filipino-English bilinguals and were given problems written either in Filipino or English. Filipino was the subjects' first language, and English was their second language. The subjects are having their mathematics education using the English language and materials written in the English language; hence they acquired the knowledge required for constructing the problem structure while being instructed in English.

As with experiment 1, by comparing the subjects' performance on the Filipino language to their performance on the English problems we can determine the effects of the language used on this important component of word problem solving. Since the knowledge required for problem-structure construction is abstract in nature, it was predicted that subjects would perform equally well in completing the problems in Filipino and in English. In other words, the language should not have an effect on the abstract elements of this component of problem solving. It was further hypothesized that the pattern of problem-completion performance would parallel solution performance. That is, subjects would also be equally successful in providing the correct solutions for problems in Filipino and in English. As with Experiment 1, this hypothesis is based on the model of word problem solving discussed earlier in which problem-structure construction is an integral part of the complete word problem-solving process. Therefore, performance in that component should have a corresponding effect on the complete task.

Method

Subjects. The participants in this experiment were 40 Grade 4 students from two public schools in Quezon City. Similar to the subjects in Experiment 1, all the students reported that they used Filipino at home, but can speak, understand, and read enough of English as it is used in their school work. The students were taught math in English, and have had substantial prior experience with arithmetic word problems.

Materials. The problems used in this experiment were directly based on the problems used in the first set of experiments. However, the questions in each of the 18 Filipino and 18 English problems were omitted. The set of 18 problems was arranged in four different random sequences. Each problem was written using big print on 5" x 8" cards. Each set of 18 problems in one random sequence was combined to make a small booklet. Each student worked on 36 problems, 18 in Filipino and 18 in English. For each student, a different random sequence was used for the Filipino problems and for the English problems. About half of the students worked on the Filipino problems first, and the rest worked on the English problems first.

Without the questions, the first two Compare problems in the Appendix became identical in form with the first two combine problems. Therefore in addition to the four easy problems identified in Experiment 1, there were two other easy problems for this experiment. That made a total of 6 easy problems and 12 difficult problems. (The set of problems in numeric format used in Experiment 1 was not used in this experiment.)

Design. Two variables were manipulated in this experiment: the language of the problem (Filipino or English) and the difficulty of the problem (easy or difficult). The two factors were studied using a 2 x 2 completely repeated factorial design. There were two dependent variables measured: the proportion of correct problem completions and the proportion of correct solutions for the problems for each of the factorial conditions.

Procedures. The student were tested individually in a reasonably quiet place in their school during school hours. All the problems were presented to the students by showing the cards that had the written problems; while the student was reading each problem, the problem was also read aloud to the student twice. Whenever possible, the student's responses were tape recorded. As with the earlier experiment, no student objected to the session being taped, although in some cases tape recording was not possible because of technical problems.

Each session began by following the same procedures described earlier to get acquainted with the student. The student was informed that they would be playing some kind of a game involving the word problems. Examples of word problems were then presented while calling particular attention to the fact that the problem had no question. The student was asked if she could think of the question that should follow. In case the student was not able to generate a question, the correct answer was given. She was told that they would be doing more of the same procedure and that for each problem she will read and generate question, she will also be asked to compute for the solution. After the student completed the Filipino and English sets of 18 problems each, she was given a small gift as a token for her participation.

Results and Discussion.

Problem Completion Data. Each of the responses for all of the problems was coded as showing correct problem completion or not. Correct problem completion was scored if the subject generated a question that was identical to or similar in structure to the omitted question. The mean proportion of the correct recall for Filipino and English, easy and difficult problems are summarized in Table 5.

The means were analyzed using a 2 x 2 Analysis of Variance (ANOVA) for completely repeated factorial design. As expected, the analysis showed a main effect of difficulty, $F(1, 39) = 141.54$, $MSe = .045$, $p.0001$; the students generated more correct questions for the easy problems.

Table 5
Mean Proportion of Correct Problem Completion (and Standard Error) as a Function of Problem Difficulty and Language of Problem in Experiment 2

	Language of Problem			
Difficulty of Problem	Filipino		English	
Easy	69.2%	(4.2)	67.0%	(3.9)
Difficult	29.8%	(4.0)	26.9%	(3.6)

However also as predicted, there was no main effect of language of the problem and there was no interaction effect either (both F's 1). These results suggest that the process of constructing the problem structure is not affected by the language of the problem.

Solution Data. The subject's computed solutions were also coded for accuracy. The mean proportion of correct solutions are summarized in Table 6. The means were also analyzed using a 2 x 2 ANOVA for completely repeated designs. Similar to the recall data, there was also a main effect of difficulty, $F(1, 39) = 124.05$, $MSe = .044$, $p<.0001$. There was also no main effect of the language of the problem, $F(1, 39)<1$; nor an interaction effect, $F(1, 39) = 3.33$, $MSe = .016$, $p=.07$. Subjects arrived at the correct solutions for the Filipino problems as often as for the English problems. Therefore, the similar levels of performance for Filipino and English problems in the problem completion results correspond to the similar levels of accuracy for Filipino and English problems. As with text comprehension results, these findings provide support for the view that since problem-structure construction is an important part of word problem solving, success in this specific component should lead to success in solving word problems. More importantly, in the process of constructing the correct problem structure and solving the problems, the subjects were not affected by the language in which the problems were written. Unlike the results of Experiment 1, the students performed just as well on the English problems, even if their proficiency in English is relatively poorer than in Filipino. These results suggest that the process of constructing the problem structure is not affected by language because the knowledge and the operations involved are abstract in nature. Therefore, at this level of processing word problems, it seems the language factors have no substantial role.

Table 6
Mean Proportion of Correct Solution (and Standard Error) as a Function of Problem Difficulty and Language of Problem in Experiment 2

	Language of Problem			
Difficulty of Problem	Filipino		English	
Easy	74.0%	(3.9)	77.1%	(4.2)
Difficult	40.8%	(4.4)	36.7%	(4.1)

Qualitative Analysis: Problem Completion and Solution Performance. The conclusion being made so far is that language factors did not affect the specific process of constructing the problem model and the overall process of generating a solution since the specific process is integral to the overall word problem-solving process. However, this conclusion was based on the parallel set of null results. To provide for more specific verification of this conclusion, the responses to the problem-completion task were recoded and compared to the solution responses. Each question provided for the problem-completion task was coded as either: a correct question (CO), an incorrect but mathematical question (MQ), or an incorrect and nonmathematical question (NQ).

A response was coded as CO if the question correctly asks for the unknown quantity regardless of the actual wording used. For example, for the problem:

Judy and Carlo have 8 candies altogether. Judy has 7 candies.

the following are CO's:

How many does Carlo have?

How many candies does Carlo have?

A total of 43.6% of the questions for the Filipino problems, and 39.4% of those for the English problems were coded as CO. It seems that there is no difference between the subjects' efficacy in constructing questions for problems in either language. A result that simply restates the earlier data.

A response was coded as MQ if the question asks about a quantity, but the quantity is not the appropriate unknown in the given problem. For example, for the above problem, the following are MQ's:

How many candies do they have altogether?

How many did Judy have?

A total of 55.7% of the questions for the Filipino problems and 59.0% of those of the English problems were coded as MQ's. Again, consistent with earlier data, these results do not show any language differences.

Finally, a response was coded as NQ, if the question did not require a quantitative response, like the following examples:

Why do they have candies?

Why does Judy have more candies?

These responses were rare. Only 0.7% and 0.1% of the responses for the Filipino and English problems, respectively, were coded an NQ's. some 1.4% of the responses for the English language was not codable in any of the categories; these were incomplete responses or non-responses.

The frequency of correct and incorrect solutions for each type of response was then coded. Table 7 summarizes these frequencies across

Table 7
Frequency of Problem Construction and Solution Performance in Experiment 2

Question Type	Solution Correct	Solution Incorrect	Total
FILIPINO PROBLEMS			
Correct Question	261	53	314
Incorrect Mathematical Question	117	284	401
Incorrect Nonmathematical Question	2	3	5
ENGLISH PROBLEMS			
Correct Question	221	63	284
Incorrect Mathematical Question	135	290	425
Incorrect Nonmathematical Question	1	0	1

Note: Frequencies are based on 18 observations for each set of Filipino and English for each of 40 subjects. Ten questions given for the English problems were not codable.

problem difficulty and across all subjects. The data show that there is a strong relationship between the accuracy or type of question given by the student and the answer that the student derives. For the Filipino problems, 83.1% of the answer for the CO questions were also correct solutions. Similarly, 77.8% of the answers for the CO questions were correct solutions for the English problems. On the other hand, 70.8% of the solutions for the MQ's for the Filipino problems were incorrect. For the English problems, the corresponding proportions was 68.2%.

The results of this qualitative analysis supports two important assertions made earlier based on the parametric analysis. First, the positive relationship between problem-completion performance and solution performance shows that success in the process of constructing the problem model led to success in deriving a solution for the problem. second, similar trends between performances in the Filipino and the English problems show that language factors did not affect problem-model construction, and also therefore, did not affect the process of deriving the problem solution.

General Discussion

In this research we investigated the effect of language factors on two specific components of understanding math word problem solving. The results of Experiment 1 showed that language proficiency is an important factor in the process of comprehending the textual information in the word problems. The results showed that subjects were better able to comprehend the text when this was in their first language or the language in which they are most proficient. On the other hand, the results of Experiment 2 showed that language seems to have no effect on the process of constructing the problem structure. The results showed that language factors do not make a difference when it comes to grasping the more abstract aspects of word problem solving in mathematics.

At the theoretical level, the significance of this study is two-fold. First, it provides empirical support for the view that there are two levels of understanding involved in word problem solving in mathematics. As stated earlier, this position has been advanced by many prominent researchers (e.g., Kintsch & Greeno, 1985; Mayer, 1985) and has been tacitly accepted in research circles. However, there has been no strong empirical verification for such a position. The experiments in this study provide an information-processing analysis of the word problem solving task. The dissociation of language effects for the two types of components described above provide strong empirical evidence for the theory.

Another theoretical contribution of this study relates to the issue of relationship between language and mathematical performance. There are different positions regarding this issue. One common wisdom is that the mathematical domain is the most abstract field of study, and hence, should not be affected by language processes. On the other, in the Philippines, some people claim that mathematics could only be properly learned in English. Still a small minority assert that math is best learned in the vernacular, Filipino. These last two positions posit some unspecified link between language and mathematical performance.

The results of the two experiments clearly show that there is a relationship between language and mathematical performance, particularly in the aspect of understanding the various pieces of information provided in the math problems. Consistent with earlier research, language processes seem to play an important role in developing efficient representations of the problem information. The results also show that the effects of language processes on the formation of problem representations lead to corresponding effects on the overall problem-solving performance. Therefore, it would seem unwise to ignore such language effects, as well as the role of language processing in mathematical performance. However, as stated earlier, understanding the problem information and the process of developing problem representations in math, at least for word problems, does not simply involve one process. There are two forms of understanding and two types of problem representations that need to be formed. Furthermore, the effects of language use are different for each level of mathematical understanding.

In the Philippines, the issue of the relationship between language use and cognitive performance is particularly significant because of the dominantly bilingual and even multilingual population. This issue is crystallized in the national debate on the medium of instruction for our educational institutions. The long standing issue and debate has not come close to a resolution inspite of constitutional provisions and a National Bilingual Policy. In the history of this issue, the discussion has been plagued with scientifically inaccurate speculative arguments and scientifically untested anecdotal evidence that is tainted with strong emotions. Sincere research efforts that address the issue have often been criticized for methodological weaknesses (see e.g., del Pilar, 1990). In this study, we brought to the debate new tools for investigating the relationship between language use and intellectual performance. The information-processing analysis undertaken in this study allowed us to have a clearer view of a specific intellectual skill, and of the particular roles language plays in the execution of such a skill. The results of the study provide a new lead in investigating the consequences of language use among bilinguals. The results suggest that we cannot resort to simplifications about the relationship between language and cognitive performance, and that there is no unitary relationship between the two. For example, when it comes to the component of textual understanding, language proficiency seems to be a critical factor; and hence, for bilinguals, using their first language would yield better results. On the other hand, when it comes to the abstract components of word problem solving, language factors seem to have no effect. To these results we can add the data of earlier studies (Bernardo, 1994a) that showed that when it comes to the process of transferring information from analogous word problems, the important language factor seems to be consistent in language use. Therefore, we should look at cognitive performance as having many specific components and that language will have a different specific function in each of these components. Any claim about the relationship between bilingualism, language use and cognitive performance should be evaluated in this light.

Based on this discussion, the practical implications of the results of this study become evident. Since the debate on the medium of instruction issue has often been fueled by misinformation and misconceptions, scientific investigations looking at the actual cognitive processes involved in the education process should lead to a more sober handling of the very important issue. This study is one such investigation that provides some scientific evidence that should bear on this debate. Moreover, the particular approach in the investigation allows us to make very specific claims about when the use of one language should be more advantageous than another and when language use does not make much of a difference.

For example, the results of Experiment 1 not only show that bilingual students comprehend the problem texts more proficiently in their first language, they also are more successful at solving problems in their first language. Therefore, the linguistic understanding of the word problems is an important ingredient in the process of learning how to solve word problems in math. In beginning to instruct children about word problem solving in math, we then need to consider which language the students will best be equipped to deal with the textual component of the problem. However, when the more abstract components of word problem solving have been acquired, we could expect that students would not be affected as much by the language in which the problems are written.

Overall, however, we should realize that the picture of the relationship between language and math problem solving is still incomplete. The study addressed the specific aspect of understanding math problems. Further studies have to be undertaken to investigate the role of language in other aspects of mathematical problem solving. With more efforts following strategies similar to that used in this research, we should soon accumulate scientific knowledge about the relationship between language and all the specific components of mathematical problem solving that will complement existing research about the effectiveness of using Filipino and English in our classrooms.

REFERENCES

Bernardo, A.B.I. (1994a). Language use and transfer in mathematical problem solving: Revisiting the Whorfian hypothesis and apprising the medium of instruction issue. *Philippine Technology Journal, 19*(1), 37-46.

Bernardo, A.B.I. (1994b). Problem-specific information and the development of problem-type schemata. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 379-395.

Bernardo, A.B.I. & Okagaki, L. (1994). Roles of symbolic knowledge and problem-information context in solving word problems. *Journal of Educational Psychology, 86*, 212-220.

Briars, D.J. & Larkin, J.H. (1984). An integral model of skill in solving elementary word problems. *Cognition and Instruction, 1*, 245-296.

Cummins, D.D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20*, 405-438.

Davis-Dorsey, J., Ross, S.M., & Morrison, G.R. (1991). The role of rewording and context personalization in the solving of mathematical word problems. *Journal of Educational Psychology, 83*, 61-68.

De Corte, E. & Verschaffel, L. (1987). The effect of semantic structure on first-graders' strategies for solving addition and subtraction word problems. *Journal for Research in Mathematics Education, 18*, 363-381.

De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representation and solutions. *Journal of Educational Psychology, 77*, 460-470.

De Corte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of Educational Psychology, 82*, 359-365.

Department of Education and Culture. (1976). *Report of the survey of outcomes of elementary education*. Makati: EDPITAF.

Del Pilar, G.H. (1990). *Isang rebista ng mga empirical na pag-aaral sa pinagkumparang bisa ng Ingles at wikang Pilipino bilang midyum ng pagtuturo*. Unpublished manuscript, University of the Philippines, Diliman, Q.C.

Kintsch, W. & Greeno, J.G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*, 109-129.

Krutetskii, V.A. (1976). *The psychology of mathematical abilities in schoolchildren*. Chicago, Chicago University Press.

Mayer, R.E. (1985). Mathematical abilities. In R.J. Stenberg (Ed.), *Human abilities: An information-processing approach* (pp. 127-150). NY: Freeman.

Nesher, P. & Katriel, T. (1977). A semantic analysis of addition and subtraction word problems in arithmetic. *Educational Studies in Mathemctics, 8*, 251-269.

Nesher, P. & Teubal, E. (1975). Verbal cues as an interfering factor in verbal problem solving. *Educational Studies in Mathematics, 6*, 41-51.

Novick, L.R. & Holyoak, K.J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 398-415.

Reed, S.K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 106-125.

Riley, M.S., & Greeno, J.G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction, 5*, 49-101.

Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem-solving ability in arithmetic. In H.P. Ginsburg (Ed.), *The development of mathematical thinking* (pp.153-196). New York: Academic Press.

Ross, B.H. (1984). Remindings and their effects in learning cognitive skills. *Cognitive Psychology, 16*, 371-416.

Ross, B.H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 629-639.

Ross, B.H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 456-468.

Ross, B.H. & Kennedy, P.T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 42-55.

Schoenfeld, A.H. (1985). *Mathematical problem solving*. San Diego, CA: Academic Press.

Sternberg, R.J. (1977). *Intelligence, information processing, and analogical reasoning*. Hillsdale, NJ: Erlbaum.

Sternberg, R.J. (1985). What is an information-processing approach to human abilities. In R.J. Sternberg (Ed.), *Human abilities: An information-processing approach* (pp. 1-4). New York: Freeman.

Appendix

Complete list of word problems in English

Combine problems

(A) Judy has 3 candies. Carlo has 5 candies. How many candies do they have altogether.

(B) Judy and Carlo have some candies altogether. Judy has 2 candies. Carlo has 4 candies. How many candies do they have altogether?

(C) Judy has 4 candies. Carlo has some candies. They have 7 candies altogether. How many candies does Carlo have?

(D) Judy has some candies. Carlo has 6 candies. They have 9 candies altogether. How many candies does Judy have?

(E) Judy and Carlo have 8 candies altogether. Judy has 7 candies. How many candies does Carlo have?

(F) Judy and Carlo have 4 candies altogether. Judy has some candies. Carlo has 3 candies. How many does Judy have?

Change problems

(G) Judy had 3 candies. Then Carlo gave her 5 candies. How many candies does Judy have now?

H) Judy had 6 candies. Then she gave 4 candies to Carlo. How many candies does Judy have now?

(I) Judy had 2 candies. Then Carlo gave her some candies. Now Judy has 9 candies. How many candies did Carlo give to her?

(J) Judy has 8 candies. Then she gave some to Carlo. Now Judy has 3 candies. How many candies did she give to Carlo?

(K) Judy had some candies. Then she gave Carlo her 3 candies. Now Judy has 5 candies. How many candies did Judy have in the beginning?

(L) Judy had some candies. Then she gave 2 candies to Carlo. Now Judy has 6 candies. How many candies did she have in the beginning?

Compare Problems

(M) Judy has 5 candies. Carlo has 8 candies. How many candies does Carlo have more than Judy?

(N) Judy has 6 candies. Carlo has 2 candies. How many candies does Carlo have less than Judy?

(O) Judy has 3 candies. Carlo has 4 candies more than Judy. How many candies does Carlo have?

(P) Judy has 5 candies. Carlo has 3 candies less than Judy. How many candies does Carlo have?

(Q) Judy has 9 candies. She has 4 candies more than Carlo. How many candies does Carlo have?

(R) Judy has 4 candies. She has 3 candies less than Carlo. How many candies does Carlo have?

Complete list of word problems in Filipino

Combine problems

(A) Si Judy ay may 3 kendi. Si Carlo ay may 5 kendi. Ilan lahat ang kendi nila kapag ipinagsama?

(B) Sina Judy at Carlo ay mayroong mga kendi. Si Judy ay mayroong 2, 4 din ang kay Carlo. Ilan lahat ang kendi nila kapag ipinagsama

(C) Si Judy ay may 4 na kendi. Si Carlo ay mayroon ding kendi. 7 lahat ang mga kendi nila kapag ipinagsama-sama. Ilan ang kendi ni Carlo?

(D) Si Judy ay mayroong mga kendi. Si Carlo ay may 6 na kendi. 9 lahat ang mga kendi nila kapag ipinagsama-sama. Ilan ang kendi ni Judy?

(E) Kung pagsasamahin, mayroong 8 kendi sina Judy at Carlo. Kung 7 sa mga kendi ang kay Judy, ilan ang kendi ni Carlo?

(F) Kung pagsasamahin, mayroong 4 na kendi sina Judy at Carlo. Kung 3 sa mga kendi and kay Carlo, ilang kendi ang kay Judy?

Compare problems

(G) Si Judy ay mayroon nang 3 kendi. Binigyan pa siya ni Carlo ng 5. Ilang kendi na ngayon mayroon si Judy?

(H) Si Judy ay mayroong 6 na kendi. Ibinigay niya kay Carlo ang 4. Ilang kendi na lang ang natitira kay Judy?

(I) Si Judy ay mayroong 2 kendi. Binigyan pa siya ni Carlo ng mga kendi. 9 na ngayon ang kendi ni Judy. Ilang kendi ang ibinigay ni Carlo kay Judy?

(J) Si Judy ay mayroong 8 kendi. Binigyan niya ng kaunti si Carlo. Ngayon, 3 na lang ang natirang kendi kay Judy. Ilang kendi ang ibinigay ni Judy kay Carlo?

(K) Si Judy ay mayroong mga kendi. Binigyan niya ng 3 kendi si Carlo. Ngayon, 5 ang kendi ni Judy. Ilang kendi mayroon si Judy bago niya binigyan si Carlo?

(L) Si Judy ay mayroong mga kendi. Binigyan niya ng 2 kendi si Carlo. Ngayon, 6 ang kendi ni Judy. Ilan lahat ang kendi ni Judy sa simula?

Compare problems

(M) Si Judy ay mayroong 5 kendi, 8 naman ang kay Carlo. Ilang kendi and lamang ni Carlo kay Judy?

(N) Si Judy ay mayroong 6 na kendi. 2 naman ang kay Carlo. Mas kaunti ng ilan ang kendi ni Carlo kaysa kay Judy?

(O) Si Judy ay mayroong 3 kendi. Mas marami nang 4 ang kendi ni Carlo kaysa kay Judy. Ilang kendi mayroon si Carlo?

(P) Si Judy ay mayroong 5 kendi. Mas kaunti ng 3 ang mga kendi ni Carlo kaysa kay Judy. Ilang kendi mayroon si Carlo?

(Q) Si Judy ay mayroong 9 na kendi. Mas marami nang 4 ang kendi niya kaysa kay Carlo. Ilang kendi mayroon si Carlo?

(R) Si Judy ay mayroong 4 na kendi. Mas kaunti ng 3 ang kendi niya kaysa kay Carlo. Ilang kendi mayroon si Carlo?

USING TEST DATA TO ASSESS DIVISIONAL AND REGIONAL PERFORMANCE OF PHILIPPINE SCHOOLS*

Abraham I. Felipe
EDUVISION 2000

The present study is an attempt to assess certain aspects of an educational system using two indices developed from test data. In the traditional way that tests have been used for assessment, groups are compared on test statistics. In the present approach, a criterion score on a test is fixed, and population samples are subsequently differentiated in terms of how they meet the fixed criterion score. The two indices were: (a) a measure of density of a particular event in a given population, and (b) a measure of concentration of the event around an expected value. It is believed that the method is more suited to a case wherein one wishes to talk about a population being assessed. The other advantage is that the indices are constructed as ratio scale. The study reports the outcomes of several exercises in which the indices were used to assess the performance of administrative units and certain types of schools in the Philippines.

The evaluation of a country's educational system is a valid and important concern. Since the late 70's, two comprehensive evaluations had been undertaken for the purpose of guiding educational reforms. In addition were other studies with less comprehensive coverages to evaluate more limited aspects of the Philippine system of education. Those studies that focused on student learning made heavy use of test data. There had been no study in which test data were used to assess other aspects of the system. The reason for this seems to be the historical use of tests in education: tests have been developed to measure what students learn or are prepared to learn. This use of tests, however, is quite limited. Test data could be used for other purposes. The present study is an attempt to show one such use.

The National Elementary Assessment Test (NEAT) and the National Secondary Assessment Test (NSAT) were adopted by the Department of Education, Culture and Sports (DECS) in 1993 and 1994 respectively, in order to assess student learning at the elementary and secondary levels, respectively. As an expression of the trust of DECS on what these tests measure, a portion of the final grades of graduating students was derived from the students' test results. Nothing else was made explicit about the other values of the tests. Other uses of the tests could only be inferred.

---

* *Paper read at the 2nd Asia-Pacific Conference on Educational Assessment, 4-8 September 1995, Chiang-Mai, Thailand. Data for the study were provided by the DECS-OPS Research and Statistics Division and Dr. Lucila Tibigar, Executive Director of the National Educational Testing and Research Center.*

If the tests really measure student learning as intended, then they should be useful also for other purposes. When the DECS announced the results of the tests "in order to guide parents, teachers and school administrators", there was an explicit recognition that the tests could give guidance. What guidance they give has not been made clear. The present study is an attempt to explore some forms of guidance that the tests provide.

In the traditional way that tests have been used for assessment, groups are compared on test scores (central tendencies, dispersions, and other test statistics). The outcome of this method of using tests is to see variations among groups in their test scores.

Alternatively, one could fix a certain criterion based on test scores (e.g., a certain level on the test) and then proceed to find out how well samples in a population could meet the criterion. The outcome of this method of using tests is to see how population samples are differentiated. The present study will follow this method. It is believed that this method is more suited to a case wherein one wishes to talk about a population to be assessed. It complements the traditional use of tests in assessments. The present study, together with a limited model for making decisions, will illustrate this use.

The opportunity to apply the recommended approach was provided by the DECS announcement of the top 250 schools on the NEAT and the NSAT. Strictly speaking, *the top 250 schools* did not mean a criterion score. The *top 250* only implied a criterion score, namely, that (mean) score which differentiated the top 250 from all the others. The research question now is: what did that score say about the DECS school divisions and regions?

METHOD

Sources of data on the top 250 schools on the NEAT and NSAT were the *Manila Bulletin* and the *Philippine Daily Inquirer* issues of June 3 and 4, 1995, containing the DECS report of the results of the 1994 NEAT and NSAT administration. *Top* was determined by DECS based on a ranking of mean scores of the students from each school. The newspapers printed also the division and the region from where each of the top schools came.

For purposes of the present study, the grouping by DECS of schools in terms of the number of examinees, or *clusters*, was ignored, making it possible to have larger frequencies of *top* schools per division or per region.

In order to use the data on schools for assessing the DECS divisions and regions, the frequencies of the top schools were converted into two new measures about the divisions and regions. The first is a measure of the actual density **W** of top elementary or top high schools, as the case may be, within the division (or region). The label **W** is for *within* the division or region. This measure of density is an index of density that is internal to the division. This was done by dividing the number of *top* elementary schools in a division (or region, as the case may be) by its total number of elementary schools. This was separately done for NEAT and NSAT data.

The second is a measure of the actual density ***L*** of top schools coming from a division (or region) in the list of 250. The label ***L*** is for ***List***, to emphasize that this is the division (or region) density within the list of top 250. This was done by simply dividing the number of top schools from the division (or region) by 250. Again, this was done separately for NEAT and NSAT data.

Densities per se have little meaning. For example, a density with a zero value has an ambiguous meaning (such as, in our specific case, when a division fails to enlist even just one school in the top list). A zero density is problematic because it does not differentiate between a zero from a small division (which is not expected to produce one top school, from the viewpoint of probability) and the zero of a division with aa thousand schools (which is expected to produce some top schools, based on sheer numbers alone).

To differentiate between these two types of zeroes (and for other purposes), an index of *expected number of schools* was therefore developed and calculated for each division and region based on (a) the total number of schools, (b) the number of schools in a territory in question, and (c) an assumption of each school having a specified (in the present case, equal) chance to be part of the top list. Expected density in the top list is the ratio of expected number and the number of schools in the top list (i.e., 250).

To give added meaning to *density*, the difference between actual density ***L*** and expected density is taken; this difference will tell whether the density in the list is greater than or less than expected. This difference is then interpreted as a measure of concentration of top schools, relative to expectation. While densities are always positive, concentrations may be negative. A ***positive concentration*** means that the density of top schools is greater than expected. A ***negative concentration*** means that the density is less than expected.

Densities (***W*** and ***L***), expected densities and concentrations were then calculated, using data from the DECS-OPS Office of Research and Statistics on the number of elementary and high schools from each division. In 1994, there were 34,179 elementary schools and 5,606 high schools, public and private, all over the country.

Similar measures were calculated on 5 types of schools -- the public schools, the private sectarian schools, the private non-sectarian schools, the laboratory schools of state colleges and universities, and the DECS science high schools (for some studies on the NSAT).The total number of sectarian schools was derived from the total membership of the Catholic Educational Association of the Philippines (CEAP) and the Association of

Christian Schools and Colleges (ACSC) in 1994. No other association of sectarian schools was included in the aggregation. The total number of non-sectarian schools was estimated from the combined total memberships of the Philippine Association of Private Schools, Colleges and Universities (PAPSCU), the Philippine Association of Private Technical Institutions (PAPTI) and the Philippine Association of Colleges and Universities (PACU). Because of a duplication of memberships in the three associations of non-sectarian schools, the figure used as the total for the non-sectarian schools might be an over-estimate. The total number of SCUs used for calculating densities was 88. The total number of DECS science high schools was as given in DECS Order No. 69, c. 1993 (i.e., n=14).

Concentrations were tested for significance using simple tests of proportions. Results are reported in terms of z-values. The significant differences between actual and expected densities are reported. Unless specifically stated, all tests are two- tailed because of a lack of basis for predicting the direction of differences.

RESULTS

National Elementary Assessment Test (NEAT)

Observations on the DECS Regions

Only three points will be established in this section: (a) the very superior performance of NCR, in comparison with the other regions; (b) the failure of one half of the regions to meet simple statistical expectations; and (c) the superior and surprising performance of the ARMM schools.

The pertinent data are given in Table 1. The first column lists the 15 administrative regions of DECS. The second column gives the total number of top schools from each region as released by DECS and printed in the *Manila Bulletin* and the *Philippine Daily Inquirer*. The third column gives the total number of elementary schools, public and private, from each region; the figures were provided by the DECS-OPS Research and Statistics Division. The fourth column gives the actual densities ***W*** of top schools per region. Actual densities ***W*** were derived by dividing each figure in the second column by its corresponding figure in the third column. The fifth column gives the actual densities ***L*** derived by dividing each figure in column 2 by the constant 250. The sixth column gives the expected densities of top schools per region. Expected densities were estimated assuming (a) only 250 will be included in the list, (b) the top 250 will be selected from a total of 34,179 schools, (c) the number of schools per region is as given in column 3; and (d) each school in the total pool initially had an equal chance of joining the top list. In the NEAT, the value of this chance is calculated to be approximately .07%. The regions are listed according to the magnitude of their actual densities ***W***, from top to bottom.

Table 1.
Regional Performance in NEAT

REGION	Number of Schools in Top 250	Number of Schools in Region	Actual Density (*W*)	Actual Density (*L*)	Expected Density	*z*	*p*
NCR	79	939	0.084	0.316	0.027	9.26	****
ARMM	32	1,677	0.019	0.128	0.049	3.14	***
IV	49	4,580	0.011	0.196	0.134	1.87	*
VII	12	1,455	0.008	0.048	0.043	0.29	
X	14	2,640	0.005	0.056	0.077	-0.95	
XII	7	1,373	0.005	0.028	0.040	-0.75	
VIII	14	2,927	0.005	0.056	0.086	-1.29	
III	12	2,688	0.004	0.048	0.079	-1.41	
IX	6	1,911	0.003	0.024	0.056	-1.83	*
VI	9	3,138	0.003	0.036	0.092	-2.57	***
II	4	1,921	0.002	0.016	0.056	-2.42	***
I	4	2,294	0.002	0.016	0.067	-2.89	***
CAR	2	1,180	0.002	0.008	0.035	-2.06	**
V	4	2,964	0.001	0.016	0.087	-3.63	****
XI	2	2,492	0.001	0.008	0.073	-3.73	****
TOTAL	250	34,179	0.007	1.000	1.000		

* $p < .10$: ** $p < .05$; *** $p < .01$: **** $p < .001$

Source of Data: DECS-OPS Research and Statistics Division

Performance of the NCR

Not surprisingly, the NCR performed better than any other region, confirming the findings of many previous comparative studies. Of NCR's 939 schools, 79 made it to the top 250 (i.e., 8.41% of all NCR schools). This density is enough to make the NCR a class by itself. No other region comes close to it. The next region in the rank has only less than 1.91% of its schools making it to the top 250, or less than 1/4 of NCR's actual density.

The density levels of each region is also presented in Table 1. The position of the NCR compared with the other regions is very secure. In the NEAT, the density ***W*** of NCR's top schools is 4.42 times more than the second ranked region (ARMM), 7.63 times more than the third (Region IV), and 62 times more than the last (Region XI).

Performance of the Other Regions

Table 1 also shows three regions with positive concentrations of top schools, meaning, they have significantly more schools in the top 250 than expected. The case of the NCR is worth noting. The absolute number of NCR schools that made it (i.e., 79) is not the important figure; it is when this figure 79 is adjusted according to NCR's number of schools (n=939) that one sees how much truly superior is the NCR compared to the others.

Table 1 also shows that concentrations were positive only for 4 regions. For the other 10 regions, concentrations were negative. Of these 10 negative concentrations, 7 were significantly lower than expected. This could only mean that half of the regions performed below par.

Regional performance (measured by the density W of top schools) is only weakly related to regional economic development which measured by the per capita regional gross domestic product indicated in the Regional Accounts of the National Statistical Coordination Board (Spearman Rank-order $rs = .42$, $p<.10$, one- tailed).

The Surprise that is the ARMM

ARMM consists of several divisions that have long been considered as deprived, disadvantaged and underserved (DDUs). As a DDU, ARMM is not expected to perform very well educationally. However, its actual performance in the NEAT was just the opposite. ARMM was one of only three regions whose density ***L*** of top schools was greater than expected. It had the second highest actual density ***W*** of top schools among all regions. Compared with the other regions, the density of top schools within ARMM was more dense. It had about twice the density of top schools as the third ranked region (Region IV, $z= 2.28$, $r=.05$) and about 2.5 times more the density of the 4th ranked region (Region VII, $z=2.64$, $r=.01$). It follows that ARMM is significantly better than any of the other remaining regions. The density of top schools in ARMM is 24 times more than that of the least performing DECS region (Region XI).

Observations on the School Divisions

DECS has 133 school divisions: 65 are city divisions and 68 are provincial divisions. Of the 65 city divisions, 31 contributed at least one school to the top 250; the 34 others did not. Of the 68 provincial divisions, 39 contributed at least one school; 29 did not. It appears that there is no difference in the number of city and provincial divisions that contributed to the top 250 list (Chi square = .31; r = n.s.).

However, two points about the city and provincial divisions could be made: (a) the density of top schools in city divisions is greater than that of the provincial divisions; (b) the density of top schools in the NCR city divisions is greater than the non-NCR city divisions. Two additional points can be emphasized from Table 2 which contains the data about divisions: (a) on the overall, the school divisions, whether city or provincial, performed lower than is expected statistically, and (b) the ARMM performance is attributable only to some of its divisions and is not the outcome of a joint effort.

Table 2.
Performance of Divisions in NEAT

DIVISION	No. of Schools in Top 250	No. of Schools in Division	Actual Density *(W)*	Actual Density *(L)*	Expected Density	*z*	*p*
Mandaluyong City	4	24	0.167	0.016	0.001	1.89	*
Quezon City	29	216	0.134	0.116	0.006	5.26	****
Tacloban City	4	37	0.108	0.016	0.001	1.82	*
Marikina/Pasig/San Juan	10	100	0.100	0.040	0.003	2.88	***
Paranaque/Las Pinas	10	103	0.097	0.040	0.003	2.87	***
Makati	4	43	0.093	0.016	0.001	1.79	*
Cebu City	6	70	0.086	0.024	0.002	2.17	**
Manila	13	155	0.084	0.052	0.005	3.24	***
Dumaguete City	1	12	0.083	0.004	0.000	0.88	
Bacolod City	5	69	0.072	0.020	0.002	1.93	*
Lanao del Sur (Maranao)	23	320	0.072	0.092	0.009	4.29	****
Marawi City	4	62	0.065	0.016	0.002	1.69	*
Mandaue City	1	17	0.059	0.004	0.000	0.83	
Muntinlupa/Taguig/Pateros	4	70	0.057	0.016	0.002	1.65	*
Siargao	6	107	0.056	0.024	0.003	2.03	**
Dagupan City	2	44	0.045	0.008	0.001	1.11	
Iriga City	2	48	0.042	0.008	0.001	1.08	
Caloocan City	3	76	0.039	0.012	0.002	1.30	
Baguio City	2	52	0.038	0.008	0.002	1.05	
Dipolog City	1	30	0.033	0.004	0.001	0.71	
Batangas	21	636	0.033	0.084	0.019	3.35	****
Pasay City	1	32	0.031	0.004	0.001	0.69	
Laguna	12	389	0.031	0.048	0.011	2.43	***
Agusan del Norte	5	172	0.029	0.020	0.005	1.51	
Iloilo City	2	74	0.027	0.008	0.002	0.92	
Rizal	5	186	0.027	0.020	0.005	1.46	
Tawi-Tawi	4	194	0.021	0.016	0.006	1.12	
Legaspi City	1	51	0.020	0.004	0.001	0.54	
Northern Samar	7	413	0.017	0.028	0.012	1.27	
Cavite	6	380	0.016	0.024	0.011	1.10	
Ormoc City	1	78	0.013	0.004	0.002	0.34	
Cagayan de Oro City	1	80	0.013	0.004	0.002	0.33	
Iligan City	1	82	0.012	0.004	0.002	0.32	
Batangas City	1	87	0.011	0.004	0.003	0.28	
Maguindanao	5	438	0.011	0.020	0.013	0.63	
Nueva Ecija	6	656	0.009	0.024	0.019	0.37	
Biliran	1	111	0.009	0.004	0.003	0.14	
Malabon/Navotas/Valenzuela	1	120	0.008	0.004	0.004	0.09	
Sultan Kudarat	2	285	0.007	0.008	0.008	-0.04	
Davao City	2	304	0.007	0.008	0.009	-0.11	
Negros Or	2	332	0.006	0.008	0.010	-0.20	
Zamboanga del Sur	5	854	0.006	0.020	0.025	-0.38	
Marinduque	1	171	0.006	0.004	0.005	-0.17	
Bulacan	3	521	0.006	0.012	0.015	-0.31	
Misamis Or	2	368	0.005	0.008	0.011	-0.32	
Bohol	2	394	0.005	0.008	0.012	-0.40	
Camarines Norte	1	234	0.004	0.004	0.007	-0.43	
Pampanga	2	490	0.004	0.008	0.014	-0.67	
Palawan	2	515	0.004	0.008	0.015	-0.74	
Cagayan	2	622	0.003	0.008	0.018	-1.00	

Table 2.
Performance of Divisions in NEAT

DIVISION	No. of Schools in Top 250	No. of Schools in Division	Actual Density *(W)*	Actual Density *(L)*	Expected Density	*z*	*p*
La Union	1	325	0.003	0.004	0.010	-0.75	
Capiz	1	378	0.003	0.004	0.011	-0.91	
Aurora	1	413	0.002	0.004	0.012	-1.01	
Isabela	2	866	0.002	0.008	0.025	-1.52	
Tarlac	1	500	0.002	0.004	0.015	-1.24	
Western Samar	1	560	0.002	0.004	0.016	-1.38	
Negros Occ	1	648	0.002	0.004	0.019	-1.57	
Pangasinan	1	1051	0.001	0.004	0.031	-2.30	**
Abra	0	250	0.000	0.000	0.007	-1.36	
Agusan del Sur	0	372	0.000	0.000	0.011	-1.66	*
Aklan	0	296	0.000	0.000	0.009	-1.48	
Albay	0	514	0.000	0.000	0.015	-1.95	*
Angeles City	0	54	0.000	0.000	0.002	-0.63	
Antique	0	422	0.000	0.000	0.012	-1.77	*
Bago City	0	35	0.000	0.000	0.001	-0.51	
Basilan	0	249	0.000	0.000	0.007	-1.35	
Bataan	0	151	0.000	0.000	0.004	-1.05	
Batanes	0	16	0.000	0.000	0.000	-0.34	
Benguet	0	266	0.000	0.000	0.008	-1.40	
Bukidnon	0	575	0.000	0.000	0.017	-2.07	**
Butuan City	0	110	0.000	0.000	0.003	-0.90	
Cabanatuan City	0	62	0.000	0.000	0.002	-0.67	
Cadiz City	0	48	0.000	0.000	0.001	-0.59	
Calbayog City	0	139	0.000	0.000	0.004	-1.01	
Camarines Sur	0	808	0.000	0.000	0.024	-2.46	***
Camiguin	0	52	0.000	0.000	0.002	-0.62	
Catanduanes	0	217	0.000	0.000	0.006	-1.26	
Cavite City	0	22	0.000	0.000	0.001	-0.40	
Cebu	0	541	0.000	0.000	0.016	-2.01	**
Cotabato	0	598	0.000	0.000	0.017	-2.11	**
Cotabato City	0	35	0.000	0.000	0.001	-0.51	
Dapitan City	0	53	0.000	0.000	0.002	-0.62	
Davao	0	581	0.000	0.000	0.017	-2.08	**
Davao del Sur	0	366	0.000	0.000	0.011	-1.65	*
Davao Oriental	0	287	0.000	0.000	0.008	-1.45	
Eastern Samar	0	402	0.000	0.000	0.012	-1.72	*
Gen.Santos City	0	69	0.000	0.000	0.002	-0.71	
Gingoog City	0	76	0.000	0.000	0.002	-0.75	
Guimaras	0	89	0.000	0.000	0.003	-0.81	
Ifugao	0	178	0.000	0.000	0.005	-1.14	
Ilocos Norte	0	313	0.000	0.000	0.009	-1.52	
Ilocos Sur	0	464	0.000	0.000	0.014	-1.85	*
Iloilo	0	935	0.000	0.000	0.027	-2.65	***
Kalinga-Apayao	0	286	0.000	0.000	0.008	-1.45	
La Carlota City	0	24	0.000	0.000	0.001	-0.42	
Lanao Del Norte	0	311	0.000	0.000	0.009	-1.52	
Lanao del Sur II	0	305	0.000	0.000	0.009	-1.50	
Laoag City	0	39	0.000	0.000	0.001	-0.53	
Lapu-Lapu City	0	24	0.000	0.000	0.001	-0.42	
Leyte	0	884	0.000	0.000	0.026	-2.58	***

Table 2.
Performance of Divisions in NEAT

DIVISION	No. of Schools in Top 250	No. of Schools in Division	Actual Density (W)	Actual Density (L)	Expected Density	*z*	*p*
Lipa City	0	71	0.000	0.000	0.002	-0.72	
Lucena City	0	47	0.000	0.000	0.001	-0.59	
Masbate	0	557	0.000	0.000	0.016	-2.04	**
Misamis Occ	0	322	0.000	0.000	0.009	-1.54	
Mt. Province	0	148	0.000	0.000	0.004	-1.04	
Naga City	0	34	0.000	0.000	0.001	-0.50	
Nueva Vizcaya	0	280	0.000	0.000	0.008	-1.44	
Occ Mindoro	0	224	0.000	0.000	0.007	-1.28	
Olongapo City	0	36	0.000	0.000	0.001	-0.51	
Or Mindoro	0	400	0.000	0.000	0.012	-1.72	*
Ozamis City	0	104	0.000	0.000	0.003	-0.87	
Pagadian City	0	64	0.000	0.000	0.002	-0.68	
Quezon	0	780	0.000	0.000	0.023	-2.42	***
Quirino	0	137	0.000	0.000	0.004	-1.00	
Romblon	0	184	0.000	0.000	0.005	-1.16	
Roxas City	0	39	0.000	0.000	0.001	-0.53	
San Carlos City	0	58	0.000	0.000	0.002	-0.65	
San Carlos City	0	59	0.000	0.000	0.002	-0.66	
San Pablo City	0	75	0.000	0.000	0.002	-0.74	
Sarangani	0	183	0.000	0.000	0.005	-1.16	
Silay City	0	22	0.000	0.000	0.001	-0.40	
Siquijor	0	38	0.000	0.000	0.001	-0.53	
Sorosogon	0	501	0.000	0.000	0.015	-1.93	*
South Cotabato	0	324	0.000	0.000	0.009	-1.55	
Southern Leyte	0	303	0.000	0.000	0.009	-1.50	
Sulu	0	420	0.000	0.000	0.012	-1.76	*
Surigao City	0	66	0.000	0.000	0.002	-0.70	
Surigao del Norte	0	236	0.000	0.000	0.007	-1.32	
Surigao del Sur	0	378	0.000	0.000	0.011	-1.67	*
Toledo City	0	27	0.000	0.000	0.001	-0.44	
Zambales	0	218	0.000	0.000	0.006	-1.27	
Zamboanga City	0	156	0.000	0.000	0.005	-1.07	
Zamboanga del Norte	0	505	0.000	0.000	0.015	-1.94	*
TOTAL	250	34179	0.007	1.000	1.000		

* p < .10: ** p < .05; *** p < .01: **** p < .001
Source of Data: DECS-OPS Research and Statistics Division

City Versus Provincial Divisions

In absolute numbers, there were less schools in the top 250 from the city divisions: 115 versus 135 from the provincial divisions. However, by sheer number alone, there should have been more top schools from the provincial divisions than the 135 reported, because there were 30,701 schools from provincial divisions as opposed to only 3,478 city schools. The superiority of the city divisions in contributing to the top 250 is incontestable after making adjustments to these population bases (z=9.04, r= .001).

The NCR Cities Versus The Non-NCR Cities

The NCR cities are the city divisions that constitute the DECS-NCR. The non-NCR cities are all city divisions outside the NCR. Of the 65 city divisions in the country, nine come from the NCR. These 9 represent 939 schools out of a national total of 3478 city schools. Even from the DECS announcement alone, the superiority of NCR city divisions is obvious: 79 top schools came from the NCR versus 36 from other cities. However, this NCR superiority is under-stated. NCR's top 79 came from a smaller number of schools in the NCR (n=939), whereas the 36 of the non-NCR cities came from a bigger population base (n=1197). When adjustment is made to the smaller NCR population base, the NCR divisions come out irrefutably superior over the other cities (z=5.24, r= .001).

Over-all Performance of the Divisions

Of the 133 school divisions, only 36 contributed either significantly more schools to the top 250, or significantly less, than expected. The contributions of the rest did not differ from expectation. Nevertheless, they can still be classified into two categories: whether they tend to be more, or less, than expected. The sign of the z-values is the guide: a plus means a division has more schools than expected; a minus sign means less than expected. Table 2 yields 23 pluses and 74 minuses. This is interpreted to mean that the concentration of top schools in the divisions, considered together, was lower than expected (z=7.08, r=.001).

The Credit to Armm's Performance

The impressive performance of ARMM noted above was not the joint work of all the ARMM divisions. ARMM did not turn out to be a homogenous region. Some ARMM divisions did not contribute a single school to the top 250. Contribution to the top 250 was made only by only 3 divisions: Lanao del Sur (Maranao) with 23, Tawi-Tawi with 4 and Maguindanao with 5 top schools. Two other divisions, Sulu and Lanao del Sur II, like 61 other divisions all over the country, did not contribute a single top school.

Observations on Schools

In the present section, observations will be made on types of schools. While there is no intention to make observations on individual schools, this will sometimes be done as a way of illustrating certain conclusions being made about certain school types.

There is a persistent interest in the comparative quality of public, non-sectarian (mostly proprietary) and sectarian schools (see Table 3). Studies routinely report comparisons among these school types. This will also be done now. In addition, some observations will be made on (a) public schools alone, at the regional level; (b) the University-Belt type of schools, and (c) ARMM schools vis-a-vis selected elite schools in Manila.

Table 3.
Comparisons of School Types in z Values, NEAT

	Public	Private Non-Sectarian	Private Sectarian
Public			
Private, Non-Sectarian	5.2 ***		
Private, Sectarian	9.98 ***	0.35	
% of Total in Top 250	0.37	7.41	7.94

*** p < .001

Public versus Private Schools

Three types of elementary schools (public, private sectarian, and private non-sectarian) are compared in Table 3 in terms of actual densities ***L*** and expected densities of top schools, and the resulting z-values when their differences were tested for significance. Briefly, both private sectarian and private non-sectarian had significantly higher concentration of top schools than the public schools, a finding that confirms other previous reports. No statistical difference was found between the sectarian and the non-sectarian schools.

The Regions and their Public Schools

ARMM was the region with the highest actual density of public schools in the top list. Table 4 provides the actual ***W***s per region. The actual density ***W*** of ARMM's public schools was 2.23 times more than that of the NCR, 5.87 times more than that of Region III, 2.60 times more than that of Region IV, 8.64 times more than that of Region VII and 57.33 times more than that of Region VI. Testing for significance between regions was no longer attempted because the frequencies of qualifying public schools by region were very few.

For this reason, it is suggested that the present density measures for public schools at the regional level be interpreted cautiously. More reliable measures of performance of the public schools must be obtained in the future, by increasing the number of schools to be covered.

The University-Belt Type

The term *University-Belt* was adopted from common use. The University-Belt type of schools refers to the big non- sectarian schools that have dominated downtown Manila. The type is now generic and is used as well to refer to similar schools outside Metro Manila. In the published list of top 250 elementary schools, there was not even a single school belonging to the University-Belt type that was reported. However, the meaning of this absence depends on the total number of schools belonging to the type, which is not presently known.

Table 4.
The Regions and Their Public Schools, NEAT

REGION	Public	Total No. of Public Schools	Actual Density (W)	Actual Density (L)	Expected Density	z	p
ARMM	32	1,669	0.019	0.128	0.052	3.00	***
IV	31	4,201	0.007	0.124	0.131	-0.23	
NCR	3	465	0.006	0.012	0.014	-0.24	
X	13	2,545	0.005	0.052	0.079	-1.23	
XII	5	1,314	0.004	0.020	0.041	-1.36	
III	8	2,451	0.003	0.032	0.076	-2.20	**
IX	6	1,866	0.003	0.024	0.058	-1.93	*
VIII	9	2,883	0.003	0.036	0.090	-2.49	***
VII	2	1,352	0.001	0.008	0.042	-2.45	***
II	2	1,858	0.001	0.008	0.058	-3.15	***
V	3	2,863	0.001	0.012	0.089	-4.00	****
I	1	2,188	0.000	0.004	0.068	-3.90	****
VI	1	2,990	0.000	0.004	0.093	-4.74	****
CAR	0	1,136	0.000	0.000	0.035	-3.03	***
XI	0	2,353	0.000	0.000	0.073	-4.44	****
TOTAL	116	32,134	0.004	0.464	1.000		

* $p < .10$: ** $p < .05$; *** $p < .01$: **** $p < .001$
Source of Data: DECS-OPS Research and Statistics Division

ARMM Schools et. al. versus Selected Elite Schools

To appreciate better the performance of Lanao del Sur (Maranao), Tawi-Tawi and Maguindanao, comparisons are made with a few *elite* schools in Metro Manila. Ateneo was ranked 7th in the 250, but two public schools in ARMM had higher ranks. La Salle Greenhills was ranked 17th, but 3 ARMM schools were ranked higher. Xavier School, Assumption College, St. Theresa and De la Salle Alabang had ranks from 25th to 33rd, and all were outranked by 4 ARMM schools. UP Integrated High School was ranked 46th and Benedictine Abbey ranked 50th. And both were outranked by 9 ARMM schools, 3 Siargao schools and 25 unheralded public elementary schools located in different parts of the country.

National Secondary Assessment Test (NSAT)

The results of the NSAT will be studied, first, in terms of whether or not they confirm the findings in NEAT, and, secondly, in terms of additional issues they raise. Thus, the structure of the report on NEAT is retained. The basic data are also analyzed as ratios.

Observations on the DECS Regions

The pertinent data are given in Table 5. The explanations given for Table 1 also apply for Table 5. The total number of secondary schools all over the country is 5606. The probability of each school initially being one of the top 250 is approximately *p*=.4%. The regions are also listed according to their actual densities ***W*** from top to bottom, as in Table 1.

Performance of the NCR

The superior performance of the NCR in comparison with the other regions that was observed in the NEAT, is confirmed in the NSAT (Table 5). The NCR remains to be a class by itself. As a matter of fact, its performance in NSAT vis-a-vis other regions even seems better than in NEAT: while only 8.41% of its elementary schools made it to the top 250 of NEAT, 20.64% of its high schools made it to the top 250 of NSAT (i.e., 71 of its 344 high schools).

With an actual density ***W*** of .206, the NCR has 2.5 times more schools in the top list than the second ranked region (Region III) and 3.1 times more than the third ranked ARMM.

Table 5.
Regional Performance in NSAT

REGION	Number of Schools in Top 250	Number of Schools in Region	Actual Density (*W*)	Actual Density (*L*)	Expected Density	*z*	*p*
NCR	71	344	0.206	0.284	0.061	6.89	****
III	41	499	0.082	0.164	0.089	2.54	***
ARMM	8	121	0.066	0.032	0.022	0.72	
IV	43	832	0.052	0.172	0.148	0.72	
VII	13	357	0.036	0.052	0.064	-0.56	
VI	16	494	0.032	0.064	0.088	-1.02	
I	17	531	0.032	0.068	0.095	-1.09	
VIII	9	346	0.026	0.036	0.062	-1.34	
CAR	4	186	0.022	0.016	0.033	-1.24	
X	7	394	0.018	0.028	0.070	-2.20	**
XI	5	284	0.018	0.020	0.051	-1.86	*
II	4	240	0.017	0.016	0.043	-1.78	*
XII	4	266	0.015	0.016	0.047	-2.01	**
IX	3	226	0.013	0.012	0.040	-1.99	**
V	5	485	0.010	0.020	0.087	-3.35	****
TOTAL	250	5,605	0.045	1.000	1.000		

* p < .10: ** p < .05; *** p < .01: **** p < .001

Source of Data: DECS-OPS Research and Statistics Division

Performance of the Other Regions

Unlike in NEAT, another region (Region III) joined the NCR in having significantly greater concentration of top schools than expected (z=2.54, p<.01). Ten regions had concentrations of top schools which were lower than expected, but only 6 of them reached significance.

Regional performance (in terms of actual density ***W*** of top schools) is positively but mildly correlated with an index of economic development, the per capita gross regional domestic product in the Regional Accounts of the NCSB (Spearman Rank-order rs = .57, p<.025, one-tailed).

ARMM

ARMM continues to be the pleasant surprise. It only ranked third in NSAT (as opposed to being second in NEAT), and had significantly fewer top schools than the NSAT second placer (Region III, z=6.24, p<.001). However, ARMM had 1.28 times more density of top schools than the 4th ranked region (Region IV, z=6.05, p.<001) and 1.82 times more than the 5th ranked (Region VII, z=12.04, p<.001). However, ARMM's concentration of top schools did not reach a significant level.

Region III is another surprise. After ranking 8th in ***W*** among the regions in NEAT, it became second placer in NSAT. At the high school level, Region III's concentration of top schools was greater than expected. The meaning of this rebound by Region III is equivocal.

Observations on the School Divisions

Data on divisional performance in the NSAT are given in Table 6. The same explanation for Table 2 applies to Table 6. The divisions are listed from top to bottom according to their actual densities ***W***.

Of the 65 city divisions, 32 contributed to the top 250. Of the 68 provincial divisions, 36 contributed to the top 250. As in the NEAT, there is no difference in the number of city and provincial divisions that contributed to the top 250 (Chi square = 2.08, n.s.)

City vs. Provincial Divisions

The superiority of the city divisions over provincial divisions that was observed in NEAT, was confirmed in the NSAT (z=9.59, p.<001). Unlike in the NEAT, however, the city divisions even had numerically more schools in the top list (i.e., 130, versus 120 from the provincial divisions), in spite of a smaller population base (see Table 6).

The NCR Cities versus the Non-NCR Cities

The superiority of the NCR cities over the non-NCR cities that was observed in NEAT, was also confirmed in the NSAT (z=11.15, p<.001). This time, however, unlike in NEAT, the NCR divisions had numerically more schools in the top list than the non-NCR cities (71 top high schools from the NCR versus 59 from the other cities). This numerical difference,

Table 6.
Performance of Divisions in NSAT

DIVISION	No. of Schoolsin Top 250	No. of Schools in Division	Actual Density (W)	Actual Density (L)	Expected Density	z	p
Quezon City	21	30	0.700	0.084	0.005	4.34	****
Iloilo City	8	19	0.421	0.032	0.003	2.44	***
Makati	4	13	0.308	0.016	0.002	1.61	*
Lanao del Sur (Maranao)	5	17	0.294	0.020	0.003	1.78	*
Dagupan City	4	14	0.286	0.016	0.002	1.58	
Mandaluyong City	3	12	0.250	0.012	0.002	1.32	
Marikina/Pasig/San Juan	9	36	0.250	0.036	0.006	2.31	**
Mandaue City	2	9	0.222	0.008	0.002	1.04	
Paranaque/Las Pinas	9	43	0.209	0.036	0.008	2.18	**
Lucena City	1	5	0.200	0.004	0.001	0.70	
Pampanga	19	96	0.198	0.076	0.017	3.15	***
Siargao	2	11	0.182	0.008	0.002	0.96	
Cebu City	7	41	0.171	0.028	0.007	1.76	*
Cavite City	1	6	0.167	0.004	0.001	0.65	
Toledo City	2	12	0.167	0.008	0.002	0.92	
Bacolod City	5	32	0.156	0.020	0.006	1.42	
Legaspi City	2	13	0.154	0.008	0.002	0.89	
Ilocos Norte	7	48	0.146	0.028	0.009	1.63	*
Angeles City	2	14	0.143	0.008	0.002	0.85	
Manila	15	106	0.142	0.060	0.019	2.37	***
Pasay City	2	15	0.133	0.008	0.003	0.82	
Batangas City	2	16	0.125	0.008	0.003	0.78	
Marawi City	1	8	0.125	0.004	0.001	0.55	
Muntinlupa/Taguig/Pateros	4	32	0.125	0.016	0.006	1.11	
San Carlos City	3	24	0.125	0.012	0.004	0.96	
Baguio City	2	17	0.118	0.008	0.003	0.75	
Laguna	15	134	0.112	0.060	0.024	2.02	**
Bataan	4	36	0.111	0.016	0.006	1.02	
Tacloban City	1	9	0.111	0.004	0.002	0.51	
Rizal	5	48	0.104	0.020	0.009	1.08	
Cagayan de Oro City	4	39	0.103	0.016	0.007	0.95	
Caloocan City	3	33	0.091	0.012	0.006	0.73	
Laoag City	1	11	0.091	0.004	0.002	0.42	
Lipa City	2	22	0.091	0.008	0.004	0.59	
Northern Samar	5	56	0.089	0.020	0.010	0.92	
Tawi-Tawi	2	23	0.087	0.008	0.004	0.56	
Pagadian City	1	12	0.083	0.004	0.002	0.38	
Quirino	1	12	0.083	0.004	0.002	0.38	
Cabanatuan City	1	14	0.071	0.004	0.002	0.30	
Bulacan	7	100	0.070	0.028	0.018	0.76	
Zamboanga City	2	30	0.067	0.008	0.005	0.36	
Cavite	5	78	0.064	0.020	0.014	0.53	
Sultan Kudarat	3	47	0.064	0.012	0.008	0.40	
Naga City	1	17	0.059	0.004	0.003	0.18	
Palawan	4	69	0.058	0.016	0.012	0.35	
Gen.Santos City	1	18	0.056	0.004	0.003	0.15	
Ifugao	1	19	0.053	0.004	0.003	0.11	
Tarlac	4	81	0.049	0.016	0.014	0.14	
Davao City	3	63	0.048	0.012	0.011	0.08	
Malabon/Navotas/Valenzuela	1	24	0.042	0.004	0.004	-0.05	

Table 6.
Performance of Divisions in NSAT

DIVISION	No. of Schools in Top 250	No. of Schools in Division	Actual Density *(W)*	Actual Density *(L)*	Expected Density	*z*	*p*
Aurora	1	25	0.040	0.004	0.004	-0.08	
Nueva Ecija	3	96	0.031	0.012	0.017	-0.48	
Isabela	3	100	0.030	0.012	0.018	-0.54	
Quezon	4	139	0.029	0.016	0.025	-0.70	
Catanduanes	1	35	0.029	0.004	0.006	-0.35	
Kalinga-Apayao	1	38	0.026	0.004	0.007	-0.42	
Western Samar	1	41	0.024	0.004	0.007	-0.49	
Marinduque	1	43	0.023	0.004	0.008	-0.54	
Surigao del Sur	1	43	0.023	0.004	0.008	-0.54	
Zambales	1	47	0.021	0.004	0.008	-0.63	
Bohol	2	98	0.020	0.008	0.017	-0.95	
Leyte	2	114	0.018	0.008	0.020	-1.17	
Negros Occ	2	114	0.018	0.008	0.020	-1.17	
Aklan	1	59	0.017	0.004	0.011	-0.86	
Batangas	2	133	0.015	0.008	0.024	-1.41	
Bukidnon	1	73	0.014	0.004	0.013	-1.10	
Albay	1	78	0.013	0.004	0.014	-1.18	
La Union	1	78	0.013	0.004	0.014	-1.18	
Cotabato	1	99	0.010	0.004	0.018	-1.48	
Pangasinan	1	278	0.004	0.004	0.050	-3.19	***
Abra	0	52	0.000	0.000	0.009	-1.53	
Agusan del Norte	0	30	0.000	0.000	0.005	-1.16	
Agusan del Sur	0	43	0.000	0.000	0.008	-1.39	
Antique	0	57	0.000	0.000	0.010	-1.60	*
Bago City	0	2	0.000	0.000	0.000	-0.30	
Basilan	0	22	0.000	0.000	0.004	-0.99	
Batanes	0	7	0.000	0.000	0.001	-0.56	
Benguet	0	36	0.000	0.000	0.006	-1.27	
Biliran	0	16	0.000	0.000	0.003	-0.85	
Butuan City	0	28	0.000	0.000	0.005	-1.12	
Cadiz City	0	8	0.000	0.000	0.001	-0.60	
Cagayan	0	92	0.000	0.000	0.016	-2.04	**
Calbayog City	0	9	0.000	0.000	0.002	-0.63	
Camarines Norte	0	41	0.000	0.000	0.007	-1.36	
Camarines Sur	0	143	0.000	0.000	0.026	-2.56	***
Camiguin	0	13	0.000	0.000	0.002	-0.76	
Capiz	0	39	0.000	0.000	0.007	-1.32	
Cebu	0	136	0.000	0.000	0.024	-2.49	***
Cotabato City	0	8	0.000	0.000	0.001	-0.60	
Dapitan City	0	8	0.000	0.000	0.001	-0.60	
Davao	0	88	0.000	0.000	0.016	-2.00	**
Davao del Sur	0	33	0.000	0.000	0.006	-1.22	
Davao Oriental	0	39	0.000	0.000	0.007	-1.32	
Dipolog City	0	7	0.000	0.000	0.001	-0.56	
Dumaguete City	0	9	0.000	0.000	0.002	-0.63	
Eastern Samar	0	41	0.000	0.000	0.007	-1.36	
Gingoog City	0	9	0.000	0.000	0.002	-0.63	
Guimaras	0	14	0.000	0.000	0.002	-0.79	
Iligan City	0	15	0.000	0.000	0.003	-0.82	
Ilocos Sur	0	78	0.000	0.000	0.014	-1.88	*

Table 6.
Performance of Divisions in NSAT

DIVISION	No. of Schools in Top 250	No. of Schools in Division	Actual Density (W)	Actual Density (L)	Expected Density	z	p
Iloilo	0	124	0.000	0.000	0.022	-2.38	***
Iriga City	0	10	0.000	0.000	0.002	-0.67	
La Carlota City	0	2	0.000	0.000	0.000	-0.30	
Lanao Del Norte	0	24	0.000	0.000	0.004	-1.04	
Lanao del Sur II	0	18	0.000	0.000	0.003	-0.90	
Lapu-Lapu City	0	9	0.000	0.000	0.002	-0.63	
Maguindanao	0	35	0.000	0.000	0.006	-1.25	
Masbate	0	79	0.000	0.000	0.014	-1.89	*
Misamis Occ	0	24	0.000	0.000	0.004	-1.04	
Misamis Or	0	57	0.000	0.000	0.010	-1.60	*
Mt. Province	0	24	0.000	0.000	0.004	-1.04	
Negros Or	0	31	0.000	0.000	0.006	-1.18	
Nueva Vizcaya	0	29	0.000	0.000	0.005	-1.14	
Occ Mindoro	0	31	0.000	0.000	0.006	-1.18	
Olongapo City	0	15	0.000	0.000	0.003	-0.82	
Or Mindoro	0	49	0.000	0.000	0.009	-1.48	
Ormoc City	0	8	0.000	0.000	0.001	-0.60	
Ozamis City	0	19	0.000	0.000	0.003	-0.92	
Romblon	0	26	0.000	0.000	0.005	-1.08	
Roxas City	0	13	0.000	0.000	0.002	-0.76	
San Carlos City	0	8	0.000	0.000	0.001	-0.60	
San Pablo City	0	8	0.000	0.000	0.001	-0.60	
Sarangani	0	0	0.000	0.000	0.000		
Silay City	0	3	0.000	0.000	0.001	-0.37	
Siquijor	0	12	0.000	0.000	0.002	-0.73	
Sorosogon	0	69	0.000	0.000	0.012	-1.77	*
South Cotabato	0	73	0.000	0.000	0.013	-1.82	*
Southern Leyte	0	52	0.000	0.000	0.009	-1.53	
Sulu	0	20	0.000	0.000	0.004	-0.95	
Surigao City	0	9	0.000	0.000	0.002	-0.63	
Surigao del Norte	0	39	0.000	0.000	0.007	-1.32	
Zamboanga del Norte	0	42	0.000	0.000	0.007	-1.37	
Zamboanga del Sur	0	105	0.000	0.000	0.019	-2.18	**
Total	250	5,605	0.045	1.000	1.000		

* p < .10: ** p < .05; *** p < .01: **** p < .001
Source of Data: DECS-OPS Research and Statistics Division

however, still under-states the underlying academic superiority of the NCR divisions. The 71 NCR contribution came from a base of 344 NCR schools, whereas the 59 from the other cities came from a larger base of 657 schools. The test of significance, however, made the necessary adjustments in their population bases.

Over-all Performance of the Divisions

Of the 133 school divisions all over the country, the concentrations in 26 divisions differed significantly from what was expected, some higher and others lower. This number of divisions with significant amounts of concentrations is comparable to what was found in NEAT.

The concentrations of top schools in the other 107 divisions were not significant. Nonetheless, one can use other information to show how these 107 divisions really performed, relative to expectation. With the sign of the z-value of each division's test of significance again as guide (a plus sign means a higher concentration of top schools than expected; a minus sign means that a division had a lower concentration than expected), Table 6 yields 36 pluses and 71 minuses. Statistically, these figures mean that when the 107 divisions were considered together, they in fact showed significantly lower concentrations (z=4.52, p<.001). This, again, is consistent with the finding in NEAT.

A Reduction of ARMM Performance

The impressive performance of ARMM in NEAT has been partially diluted in NSAT. Whereas 3 of the 5 ARMM divisions contributed to the top NEAT list, only 2 of these ARMM divisions contributed to the top NSAT list: Lanao del Sur (Maranao) contributed 5 schools and Tawi-Tawi contributed another 2. Maguindanao which contributed 5 top schools in the NEAT did not contribute a single school in NSAT, just like Sulu and Lanao del Sur II.

Observations on Schools

Public versus Private School Issue

Five types of high schools (public, private sectarian, private non-sectarian, laboratory schools of state colleges and universities or SCUs, and the science high schools created under DECS Order No. 89, c. 1993) were compared. The results are in Table 7, showing actual and expected densities of top schools, and the resulting z-values when their differences were tested for significance. The DECS-created science high schools had very high actual densities. However, one should be careful in making a conclusion from this observation because of the very small size of the sample (n=14). Of all types of schools, it is the SCU lab schools that have the highest concentration of top schools, significantly more than any other type - private sectarian, private non-sectarian, or the regular DECS high schools. Except for the DECS science high schools, all high school types were superior to the regular DECS high schools. The DECS science high schools also tended to be superior than the public schools, but the superiority did not reach a significant level probably because of the small sample. The sectarian and the non-sectarian schools did not perform differently from each other.

An interesting case is that of schools that have the phrase *science high school* in their names (e.g., Cabanatuan City Science High School) but are not in the list of DECS science high schools. There is a very evident presence of these science high schools in the list of top 250, suggesting

Table 7.
Comparisons of School Types in z Values, NSAT

	Public	Private Non-Sectarian	Private Sectarian	SCU	Science High School
Pubic					
Private, Non-Sectarian	5.39 ***				
Private, Sectarian	11.05 ***	1.36			
SCU	6.51 ***	4.80 ***	4.47 ***		
Science High School	1.84 *	1.13	1.01	1.05	
% of Total in top 250	0.01	8.67	11.16	28.85	21.43

*** p < .001: * p < .10

that they seem to be a type worth noting. However, it is difficult to assess them as a group because there is no record of their total number. Future studies may follow up this observation.

The Regions and their Public Schools

Once again, ARMM had the highest actual density ***W*** of top public high schools. Table 8 gives the actual densities of top public high schools per region. The actual density ***W*** in ARMM is 2.67 time more than that of the NCR, 3.68 times more than that of Region III, 5.49 times more than that of Region IV, 8.30 times more than that of Region VII, and 29.18 times more than that of Region VI.

Just as in the case of NEAT, however, caution is advised in interpreting these measures of densities of public high schools in the regions because of the small samples on which they were based.

The University-Belt Type

Like in NEAT, no school of the university belt type was observed among the top 250 schools in NSAT.

ARMM and Selected Elite Schools

The Lanao schools did not perform as impressively as they did in the NEAT, in comparison with some Metro Manila *elite* schools. Unlike in the NEAT, no Lanao school had a mean score higher than Ateneo de Manila, Xavier, Assumption, Miriam, UP Integrated, La Salle Alabang, Sta. Escolastica. Only one ARMM school (Rank 74 in the Bulletin list) ranked higher in mean scores than two elite private schools in Metro Manila (Benedictine Abbey and San Beda).

Integrating Some NEAT and NSAT Results

Does NEAT mirror NSAT? The present study will address the question at the level of the DECS divisions and regions.

Table 8.
The Regions and Their Public Schools, NSAT

REGION	Number of Schools in Top 250	Number of Schools in Region	Actual Density (*W*)	Actual Density (*L*)	Expected Density	*z*	*p*
ARMM	7	69	0.101	0.101	0.012	2.41	***
NCR	3	102	0.029	0.029	0.018	0.60	
III	6	279	0.022	0.022	0.050	-1.74	*
I	7	360	0.019	0.019	0.064	-2.61	***
IV	7	478	0.015	0.015	0.085	-3.82	****
XII	2	147	0.014	0.014	0.026	-0.91	
VIII	3	304	0.010	0.010	0.054	-2.88	***
VII	2	211	0.009	0.009	0.038	-2.05	**
X	2	235	0.009	0.009	0.042	-2.38	***
II	1	144	0.007	0.007	0.026	-1.54	
XI	1	224	0.004	0.004	0.040	-2.70	***
VI	1	322	0.003	0.003	0.057	-3.61	****
CAR	0	107	0.000	0.000	0.019	-2.21	**
IX	0	163	0.000	0.000	0.029	-2.74	***
V	0	330	0.000	0.000	0.059	-3.95	****
TOTAL in top 250	42	3,475	0.012	0.242	0.620		

* p < .10: ** p < .05; *** p < .01: **** p < .001
Source of Data: DECS-OPS Research and Statistics Division

Correlation of the Regions

The actual densities *W* of top schools in NEAT are significantly correlated with the actual densities *W* in NSAT (Spearman Rank-order rs=.63, p<.01, one-tailed). This suggests that regions with good elementary schools also tended to have good high schools. This is not surprising since there should be some tendency to administratively standardize policies and practices for various educational levels within each region. It would have been more troublesome otherwise.

Table 9 summarizes how the regions performed in the two tests. Table 9 uses three categories of performance on both the NEAT and NSAT (positively significant concentration, negatively significant concentration, and not significant), allowing for a total of 9 combinations.

As earlier mentioned, only NCR had a significantly greater concentration than expected in both the NEAT and the NSAT. On the other hand, there were four regions whose concentrations of top schools were less than expected in both tests (Regions II, V, IX and XI).

The other 10 regions did not have a consistent showing in both tests. There was no case of a region which was significantly positive in one assessment test and significantly negative in the other.

Table 9.
Correlation of NEAT and NSAT Performance, REGIONAL LEVEL

		NEAT		
		SIGNIFICANTLY POSITIVE	SIGNIFICANTLY NEGATIVE	NOT SIGNIFICANT
NSAT	SIGNIFICANTLY POSITIVE	NCR		Region III
	SIGNIFICANTLY NEGATIVE		Region IX Region II Region V Region XI	Region X Region XII
	NOT SIGNIFICANT	ARMM Region IV	Region VI Region I CAR	Region VII Region VIII

Correlation of the Divisions

Table 10 is a summary of how the divisions performed in the two assessment tests. Table 10 is Table 9 carried at the level of DECS divisions. As in Table 9, entries in Table 10 were based on the behavior of each division on a statistical test, as reported in Tables 2 and 6. Divisions which failed to reach statistical significance in at least one test (whether NEAT or NSAT) were not listed anymore. Ninety-two DECS divisions are to be lost in anonymity because of this guideline.

At the divisional level, the actual density ***W*s** of top schools in NEAT are significantly correlated with the actual density ***W*s** in NSAT (Pearson r=.68, p<.001). This means, expectedly, that divisions with good elementary schools also tended to have good high schools. The expected densities in NEAT and NSAT are correlated at .85 (p<.001).

Table 10 shows 7 divisions which had higher concentration than expected in both NEAT and NSAT. Five of these 7 are NCR schools. Only 11 of the 133 divisions had the unenviable record of failing to meet expectation in both NEAT and NSAT. Twenty three divisions were not significant in one test but were significant (positively or negatively) in the other. There is also no case of a division which was significantly positive in one test and significantly negative in the other.

Performance Versus Expectation

Until this point, the term *expectation* or *expected number* has been used in a very restrictive way in order to avoid imprecision. An expectation is a value that can be estimated given the size of a sample, the size of the total population, and the probability of occurrence of an event.

Table 10.
Correlation of NEAT and NSAT, DIVISIONAL LEVEL

		NEAT		
		SIGNIFICANTLY POSITIVE	SIGNIFICANTLY NEGATIVE	NOT SIGNIFICANT
	SIGNIFICANTLY POSITIVE	Cebu City Makati Manila Lanao Mar Parañaque Quezon City Marki/Pasig		Ilocos Norte Iloilo City Laguna Pampanga
NSAT	SIGNIFICANTLY NEGATIVE		Antique Davao Cebu Iloilo Masbate Camarines Sur Pangasinan Cotabato Sorsogon Ilocos Sur Zambo del Norte	Cagayan Misamis Oriental
	NOT SIGNIFICANT	Bacolod Batangas Mandaluyong Marawi Munti/Pateros Pasay Siargao Tacloban	Agusan del Sur Albay Bukidnon Davao del Sur East Samar Leyte Oriental Mindoro Quezon Sulu Surigao del Sur	

In this section, the term *expectation* and related concepts (e.g., *expected*, *not expected*, *unexpected*) will be used more loosely, to permit the more subjective elements that are incorporated in a *guess*, a *prediction*, a *feel*, and even a *demand* or a *requirement* (as in the sentence "I *expect* you to do this").

These subjective expectations may come from many sources. They could have a part built on statistical probability. They could have parts from subjective impressions about aspects of the environment which one validates consensually whenever he gets a chance. Some of these validated impressions form what are called *reputations*. They could also have parts from some theory by which one systematically constructs and reconstructs his world (e.g., a sociological theory of development which incorporates a relation between economic and educational development, or a psychological theory incorporating a relation between home background, motivation and academic learning).

Table 11 attempts to relate the performance of the various divisions with subjective expectations. Divisions were classed into whether they were expected (or not expected) to have more (or less) schools in the top list. A division is *expected* to do well if it is an urban area, economically developed, has a large number of schools, and enjoys a reputation for having good schools. On the other hand, a division is *not expected* to do well if it is in a backward and/or undeveloped area, has few schools, and has no reputation for good school work. The measure of whether they had actually more or less schools in the top list was an objective measure, namely, having the difference between the actual and the expected densities meet the acceptable level of statistical significance. Having more than the expected number of schools in the list is positive in Table 11; having less schools negative. Only divisions with statistically significant concentrations of top schools were listed in Table 11. This criterion explains why the list is short.

Table 11.
Expectation and Concentrations

	NEAT		NSAT	
	EXPECTED	NOT EXPECTED	EXPECTED	NOT EXPECTED
HIGH CONCENTRATION	Bacolod Cebu City Makati Mandaluyong Manila Marikina Muntinlupa Parañaque Quezon City	Lanao del Sur (mar) Marawi Siargao Tacloban	Bacolod Cebu City Cagayan de Oro Dagupan Ilocos Norte Iloilo Laguna Makati Mandaluyong Manila Marikina Muntinlupa Pampanga Parañaque Quezon City Rizal	Bataan Lanao del Sur (Mar) Northern Samar
LOW CONCENTRATION	Bukidnon Camarines Cotabato Isabela Leyte Masbate	Cebu Davao Iloilo Negros Occidental Pangasinan Quezon	Cagayan Camarines Sur Zambo del Sur	Cebu Davao Iloilo Pangasinan

Performance and Expectation in NEAT

Eleven (11) divisions, of which 10 are city divisions, had significantly higher concentrations of top schools, as expected. Of these 10 city divisions, 8 are NCR divisions. Many more city divisions had schools in the top 250, but they did not have sufficient numbers to be significant. In 13 provincial divisions, concentration of top schools was significantly lower, as expected. In their case, a poor performance was predicted because of their level of economic development and the absence of a compensating reputation.

Attention is called to two types of unexpected and interesting cases. The first consists of 4 rural and underdeveloped divisions with very high concentrations: Lanao del Sur (Maranaw), Marawi, Siargao and Tacloban. The second consists of 8 very large divisions with hundreds of schools under each but which failed to send even a single school to the top list (except for Pangasinan which had 1,051 schools and had 1 school in the top 250). All these cases are shown under the column *Not Expected*.

Performance and Expectation in NSAT

Nine divisions performed positively as *Expected*. Of these 9, only 2 were provincial divisions and 7 were city divisions. The 2 provincial divisions are both around Metro Manila.

Again, the two types of *Unexpected* cases must be documented. First, the case of rural and under-developed divisions that turned out to have significantly higher concentrations of top schools. In NSAT, they were two - Lanao del Sur (Maranaw) and Ilocos Norte; in the NEAT there were four. Among all 133 divisions, only one had significantly high concentrations of top schools in both NEAT and NSAT - Lanao del Sur (Maranao), against many odds. Lanao's three companions that were able to do this in the NEAT, were unable to keep up with their performance in the NSAT.

The second type of *Unexpected* cases is again the large divisions with hundreds of schools but which failed to have a single school in the top list. By the sheer number of schools under them, they could have produced some top schools. Apparently, they had not. Of the 8 divisions that performed this way in NEAT (Bukidnon, Cebu, Cotabato, Davao, Iloilo, Ilocos Sur, Pangasinan and Quezon), five performed in the same way in NSAT (Cebu, Cotabato, Davao, Iloilo and Pangasinan).

DISCUSSION

The discussion will focus on only a few findings which raise issues that have practical (policy and/or management) implications. There are other findings of practical and some of theoretical value, that will not be touched; they are left to others who are interested in them. But first are some theoretical perspectives to lay the premises of the present study.

The Measures Used

Let us first dispose of the issue about the data used in the analyses, namely, proportion of schools making it to the top 250. The usual and traditional approach of studies involving test data is to get test scores, in raw or converted forms, and calculate for means, dispersions and other statistics, in order to characterize or differentiate groups. After these, usually, generalizations are made.

Obviously, the traditional approach has its uses. When one is interested in issues wherein magnitudes of test scores are important, the usual approach should be more appropriate. Thus, if the traditional approach was used on the NEAT and NSAT data published by DECS, it is very probable that the findings will be expressed in statements like *"Region A has significantly higher scores than Region B"*. Historically, however, this type of statements has made little practical consequences.

More satisfactory alternatives are in order. It is believed that studies that end up with statements of the form *"Region A is x% better than Region B."* will have greater practical value. Can such statements be made using test data? This is the methodological problem. The present exercise was an attempt to show a method for doing this.

The key to doing this is the kind of measure used, according to theorists. Scores in tests are, at best, interval measures. As such, what one can do with them are limited. For example, one cannot perform arithmetic operations with them. Hence, the simple question *"How much better is Region A than Region B?"* cannot be answered, because this question requires applying arithmetic operations. However, one can answer the question *"Is Region A better than Region B?"* which only requires that the two regions be ranked in some kind of order.

To answer the question *"How much better is Region A than Region B?"*, one has to use ratio scales. The actual and the expected densities of top schools in the present study are examples of ratio scales. They add meaning to observations because they convert such observations first on a common scale. In the process, they help add meaning to some generalizations about *owners* of test scores. They enable us to answer questions about how much more or how much less, because they permit the use of arithmetic operations.

Without the conversion into ratios, the actual frequencies could be misleading. The meaning for two divisions having the same number of top schools (say, 4) will depend on how many schools each has. If one has only 40, and the other 400, clearly the latter is performing worse. The conversion into proportions compensates for the impact of sheer size or numbers. A division with 1000 schools has more chance of being represented in the top list than the division with only 40 schools, based on sheer probability alone. Converting into proportions puts all on a common footing.

The search for appropriate scales led to the measures of *actual density* (***W*** and ***L***), and *expected density*, and their derivative (concentration), of top schools that were used extensively in the present study. The major findings were based on the use of these measures. Actual density ***W*** is a characteristic of the object and as such may be correlated with other variables (e.g., some development indicators, satisfaction of political constituencies, or an index of general social well-being). The measure actual L is a corresponding point of W in a ratio scale.

Hierarchy of Regions

The ranking of regions in terms of density of top schools is as follows:

In the NEAT

NCR > ARMM > Region IV > OTHERS

In the NSAT

NCR > Region III > ARMM > OTHERS

The above ranking is based on density levels given in Tables 1 and 5. Below ARMM, the other regions were grouped into *Others* because of the very small differences in their densities. Those differences are mathematically tenable, but there is no basis to believe that it will make any difference to recognize them. A rougher ranking may be sufficient, instead.

The second placer has a shaky position. First, no region is second placer for both NEAT and for NSAT. Second, the second placer is closer to the remaining regions than to the first placer, i.e., although the second place is a place of honor among peers, it is really a poor second place. Between Region III and ARMM as a second placer, ARMM seems to be more stable (it did not stray very far when it failed to get the second highest density W in the NSAT portion of the present study; it remained third).

Hierarchy of Divisions

In both the NEAT and the NSAT, the divisions may be ranked in terms of the following hierarchy:

NCR Divisions > Other City Divisions > Provincial Divisions near the NCR > Other Provincial Divisions

As noted earlier, the NCR divisions were a class by themselves. That they would have the highest densities of top schools has always been taken for granted. But until this study, there has not been an attempt to estimate their densities of top schools, so far as this writer could recall. Neither had there been an attempt to compare their densities with the densities of the other divisions.

Next to the NCR divisions are the *Other City Divisions*. They did not have a very distinguished showing in either the NEAT or the NSAT, but they were clearly superior to the provincial divisions.

For purposes of future studies, the class of divisions called *Provincial Divisions* should be further refined. The present classifications, *Provincial Divisions Around the NCR* and *Other Provincial Divisions* are not very satisfying. There must be some continuum, now not yet understood, that could explain the very wide range of performance of the provincial divisions. For example, the performance of Lanao del Sur (Maranao), Tawi-Tawi, and Siargao raises questions about how clear or dim is our understanding of these divisions. So with the zero densities of huge divisions like Pangasinan, Cebu, Davao and Iloilo.

The unexpected performances of Lanao del Sur (Maranao), Tawi-Tawi and Siargao are pregnant with other more interesting possibilities. Their performance is the virtual opposite of what development theories in sociology, education, economics, psychology and other disciplines would normally predict. The unexpected findings about these divisions might just give birth to a major discovery in the social sciences. The current orthodoxy states that school learning is a product of a yet undetermined mix of inherent intellectual endowments (genetic pool), home and environmental factors that increase motivation and drive to perform well in school, and the quality of the schooling experience brought about by teachers, facilities and the like. The present NEAT and NSAT results might contain the seed of what could revise this orthodoxy. It is significant that a major social science discovery might just be born in a place not so far from the alleged home of the Tasadays.

The unexpected findings in Lanao, Siargao, Tawi-tawi and other divisions also have implications on proper government administration. If they are confirmed and re-confirmed in other aspects of social life in those divisions, it will be time to question the place of development assistance in public administration. They hint at a possible inhibitory effect of programs of assistance on some types of development (e.g., intellectual). Consequently, they might compel us to adopt a more economic system of administering government, which will seek to enhance development by refusing to assist.

The findings also suggest a need to develop a taxonomy of the provincial divisions that could explain their emergent hierarchy in the NEAT and the NSAT. For the sake of explicitness, the emergent hierarchy is as follows:

Lanao del Sur (Maranao) Siargao Marawi Divisions near NCR	>	The Large Majority	>	Pangasinan Cebu Davao Iloilo

The four divisions to the right are the four divisions whose actual densities of top schools are significantly less than expected in both the NEAT and the NSAT. The *Large Majority* at the center is a black box representing more than 50 provincial divisions, the argument for developing a taxonomy. This black box reflects the writer's lack of understanding of what could be a more suitable classification of DECS provincial divisions. And the divisions to the left (Lanao, Siargao and Marawi on one hand, and the divisions surrounding the NCR on the other) represent two opposing trends. The southern divisions are phenomenal divisions with cognitively elite school children from culturally and economically disadvantaged environments. The divisions around the NCR, on the other hand, are expected to do well because of some osmosis of development.

Hierarchy of Schools

In the order of decreasing density of top schools, the emerging hierarchy is as follows:

In the NEAT

Private Sectarian Private Non-Sectarian	>	Public Schools

In the NSAT (excluding the DECS-created science high schools)

SCUs	>	Private Sectarian Private Non-Sectarian	>	Public Schools

It appears now that, as a class of high schools, the best high schools are the lab schools of the SCUs. Individually, they do not compare to the best performing private schools. Not being flashy, they have not caught the public's attention as to their quality. They have been taken for granted. Government policy has even consigned them to be phased out. However, they have the highest density of top performing schools. As a class, they are clearly better than the sectarian schools which just happen to be more conspicuous.

In contrast are the private sectarian schools many of which had enjoyed the reputation of being good schools. The individual performance of some sectarian schools must have been responsible for this reputation. The density *L* of sectarian schools as a group, however, did not justify their reputation; their density did not significantly differ from those of the private non-sectarian schools and the DECS science high schools, whereas they had a significantly lower density than the SCUs.

The private non-sectarian schools have been the most maligned of all schools. However, they performed creditably, as good as the private sectarian and the DECS science schools, and significantly better than the public schools. However, the *University-Belt* type is conspicuously absent in the list.

Lumping together all public schools in the above schema is misleading. Many public schools, both elementary and high schools, were in the top 250 list. This means that there are public schools among the very best. However, there are 32,134 public schools altogether. What is known about them is almost nil. The present study allowed just a brief glimpse into some of these schools. We are, as it were, looking at the world of public schools through a glass darkly.

At least two things could be immediately done about the public schools. First, refine their classification. Perhaps classifying them in terms of their history: barangay, municipal, provincial, comprehensive high schools, etc. - might throw some light into what seems to be their bewildering performance. Second, vary the criteria for selecting schools to be studied. This will lead to producing different types of density measures about the schools. One way is to systematically expand the sample size by a progressive variation of the criterion of sample selection. For example, the present study may be replicated for the top 500, then the top 750, then the top 1000, etc.. Another way is to set criteria using socio-demographic variables, or curricular variations, etc. Density measures may be correlated with other variables. These are samples of future studies that could be done. It is only argued that to have a stable picture of our educational system, density (measures that are sensitive to population features) should be built on a firm foundation of non- shifting data bases.

In addition is the need to study the so-called science high schools. DECS Order No. 89, s. 1993 designated 14 high schools as *science high schools*. Three of these 14 entered the top list. In addition, there were many public schools in the top list with the phrase *science high school* in their names but were not listed in the DECS Order. These other science high schools have very impressive performance. However, we could not report on how well they performed as a group because we do not have their total number, like the case of the University-Belt type.

Future studies must also seek explanations of the significant findings in the present study which is only descriptive.

The Tyson-Douglas Paradigm: A Limited Model for Some Management Actions

Part of the interest in the present study is to develop concepts and approaches that have practical consequences in the sector of education. We believe that the concepts of actual density, expected density and concentration (say, of top schools) are some such concepts. They give more quantitatively meaningful information. Their measurements may be manipulated arithmetically. The concept of *density* has a place of honor in the field of testing, as it is implicit in the process of developing norms that are demanded in standardized tests. It is involved wherever there is a concept of *standards* based on characteristics of some populations.

The only problem is that very often the basis for evolving a measure of density is not available. *Density* requires carefully prepared, respectable and solidly measured background data, as well as valid and reliable measurements of certain observations whose densities we are interested in establishing. The field of educational testing which inspired NEAT and NSAT has emphasized care in the measurement of student performance in schools. We call attention to an equal need for background data that could produce measures of densities, in order to expand the value of test measurements being taken.

To organize the present findings and illustrate the use of the concept of density for management decisions about different aspects of the system of education, consider a simple paradigm involving two ways of classifying the outcomes of an event (say, a decision, or, in the model, the outcomes of an uneven boxing match such as the Mike Tyson vs. Buster Douglas fight): first, in terms of whether the outcome is *expected* or *not expected*; and second, in terms of whether the outcome is positive or negative. The term *expected* is not used in the mathematical sense, but in terms of being *foreseeable* and *understandable*. Being positive or negative refers to the effect being *good* or *bad*. This model was used earlier to classify findings in the present study.

Let us apply the paradigm to the results of our study on NEAT and NSAT. Let us specify reasonable courses of action to be done under the four resulting combinations in the paradigm.

Case #1. Outcome is expected and positive. This is the case of the undefeated champion. Tyson won. Example: the case of the NCR. That it would be the leading region is expected; to have its leadership confirmed was good for it.

The expected reaction of an observer to Case # 1 school is "*Siyempre naman!*".

The demand of the champion is plain recognition and respect. The champion does not need consolation, comfort, support, etc.

The appropriate treatment for Case #1 teachers. Acknowledge they are good; source them when in need for *good* teachers. How about any special incentives? The economic answer is, what for? What happened was expected. No big deal. Performing well is their responsibility, or even their job.

The appropriate treatment for Case #1 administrators. Same as for teachers. A slight nod; acknowledgment.

The appropriate treatment for Case #1 schools. Let them alone. They are probably best left alone.

Case #2. Outcome is expected but negative. This is the case of the underdog who lost. Douglas lost. Examples: the cases of Camarines, Sorsogon, Leyte. They were not expected to perform well and, in fact, they did not. This result was not good for them.

The expected reaction of an observer to Case #2 school is: "*Talagang ganiyan*", "*Wala tayong magagawa*", "*That's life*".

The need of the losing underdog is to be consoled or encouraged; perhaps he should be encouraged to resign to his fate.

The appropriate treatment for Case # 2 teachers. Console. Support. Assist.

The appropriate treatment for Case #2 administrators. Same as for the teachers.

The appropriate response to Case #2 schools. Same as for teachers. If they have been receiving help but the help had not made a difference, perhaps close the school.

Case #3. Outcome is not expected and positive. This is the case of the pleasant surprise. The case of Buster Douglas knocking out Mike Tyson. The giant killer. Or the winner of a lotto draw. Examples: the cases of Lanao del Sur (Maranao), Marawi, Siargao, Tacloban. At the level of a school, an example could be the case of Lilod Raya Elementary School (Lanao) having a higher mean score than Ateneo de Manila. These schools were not expected to perform as well as they did, but they did anyway. The result was of course good for them.

The appropriate reaction an observer to Case #3 school could be the doubting "*Talaga ba?*", the American imperative "*Oh, come on now.*", "*Are you pulling my leg?*", "*Is he putting me on?*" or the innocent exclamation "*Ang suwerte naman!*".

The demand of the winning underdog is respect which one may only withhold at his own peril.

The appropriate treatment for Case #3 teachers. Acknowledge them for outstanding teaching achievements. Source them for teachers to be used as models or resource persons for how to overcome odds or to revise existing theories relating development to school achievement.

The appropriate treatment for Case #3 administrators. A plaque or even a statuette plus a promotion should be the least. If articulate, recommend for lectureship within the country and abroad on how to fight and overcome the vicissitudes of poverty and lack of opportunities. Recommend for major awards, e.g., the Magsaysay Award.

The appropriate treatment for Case #3 schools. Make as a showcase for international educational accomplishment. Invite international visitors and scholars to see what can still be done under very hostile circumstances.

Case #4. Outcome is not expected and negative. The case of the champion who lost the fight. The outcome was not expected, but the champion was knocked out. Examples: the cases of Cebu, Pangasinan, Iloilo, Davao, Quezon. Nobody expected that with hundreds of schools under each, they could not have even a single top school. Of course, the result was disappointing (negative) for them.

The appropriate reaction of an observer to Case #4 school is: "*Anong nangyari?*", "*What happened?*", Or, the disappointed "*Hindi pala sikat!*". Or, the innocent "*Malas!*".

The wish of the losing champion is for him not to be noticed by others; an alternative is to protest.

The appropriate treatment for the Case #4 teachers is to ask them to explain.

The appropriate treatment for the Case #4 administrator is also to ask him to explain; after asking him to explain the first time, ask him to explain a second time so he does not miss the point that people are concerned with what had happened. The extreme treatment is to change him.

The appropriate treatment for the Case #4 school is .... what? What can you do a school when its students perform miserably? Is the school, as a school, at fault? Maybe, just change its head.

A reader may disagree with the recommended courses of action under the different contingencies in the model. After all, different schools of management thought proceed from different assumptions and move to different directions. However, the point that density measures have practical uses has been amply illustrated. Although the model was illustrated using schools as examples, it obviously can be applied as well to divisions and regions.

Some findings deserve to be public topics. They raise many intersecting issues in the fields of psychometrics and education, government, politics and ethics. For many findings to be accepted, it is legitimate to raise questions about the tests' ability to meet the basic psychometric standards of reliability and validity, and the usual standards of administration and scoring. Many findings challenge the current orthodoxy that requires economic development as a condition for social development. They raise issues on the wisdom of assistance programs for underdeveloped communities. There is also a finding that is pertinent to the government policy to phase out the high schools of the SCUs that turned out to be the best class of high schools in our system. A concentration of cognitive endowments in some disadvantaged communities, along boundaries that are geo-political, that does not radiate to nearby geographic areas, should be checked and re-checked for authenticity since that finding amounts to a major discovery in the social sciences which, if alleged and then disconfirmed later, will embarass the sources of these data no end. DECS adopted NEAT and NSAT in order to make more information available to students, parents and other consumers of education in the process of their choosing schools. Therefore, it has a responsibility to ensure that the data it gives the public on NEAT and NSAT actually inform and guide, rather than mislead. These and many other issues ought to be discussed candidly, without fear of doing something that is not politically correct.

Finally, attention is called to the present use of test data for assessing non-learning outcomes. In the present exercise, test data were not used to monitor learning but, instead, to describe and differentiate divisions, regions and types of schools. That method of using test data is new and, with the present study, can now be evaluated as a method in educational assessment. In the future, more care will be needed in using NEAT and NSAT data from new test administrations, since the system of managing the first NEAT and NSAT might corrupt the validity of the tests as instruments for assessing learning.