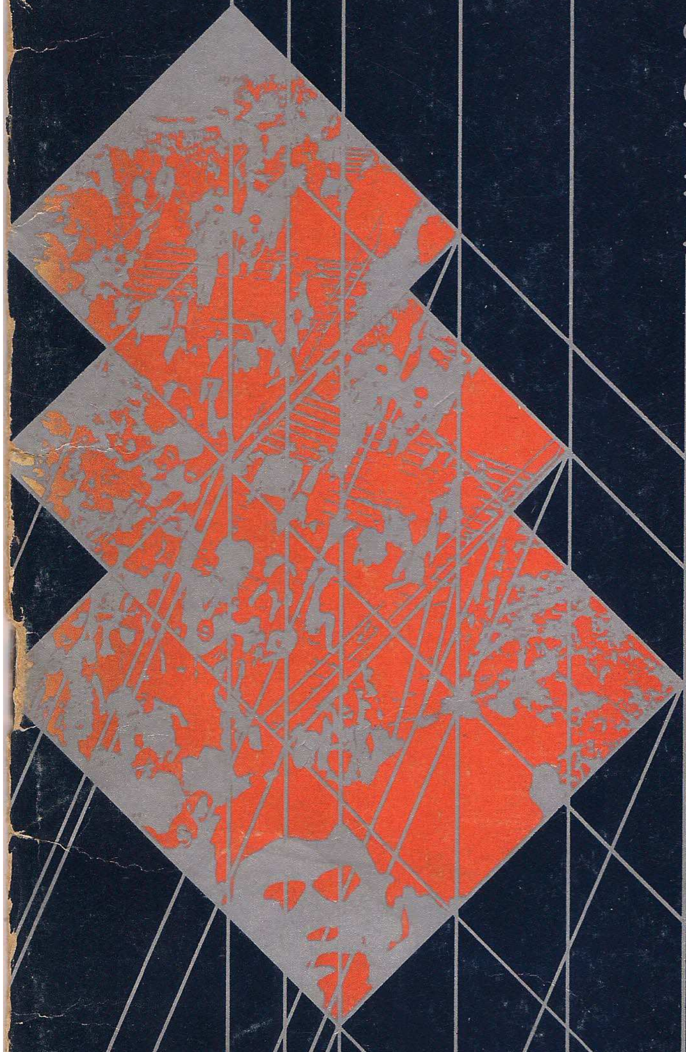
Published by the
CENTER FOR
EDUCATIONAL
MEASUREMENT

Philippine Journal of Educational Measurement

First Issue • September 1979



INAUGURAL NOTE

The issue begins CEM's *Philippine Journal of Educational Measurement.* With this comes our first effort to provide our member schools with the professional and technical issues that are relevant considerations in educational testing. The idea is to inform, stimulate, and challenge perceptions about the reason, process and consequence of tests.

At no other point in the history of education in the Philippines has testing been at once the subject of praise and doubt. The praise borders on the excessive expectations that testing's advocates have for what tests *can* do. The doubt stems from the perceptions people have about the ability of tests to do what they have been *promised* to do.

This publication is meant to distill the essence of what is current and helpful in testing. In this regard, the *Journal* must take the lead by getting into current issues in measurement in the spirit of disinterested commitment and a willingness to stimulate public discussion of schools' testing concerns.

ABRAHAM I. FELIPE, Ph.D.
President
Center for Educational Measurement

PHILIPPINE JOURNAL
of
EDUCATIONAL MEASUREMENT

Member, Publishers Association
of the Philippines, Inc.

SEPTEMBER 1979

FIRST ISSUE

The *Philippine Journal of Educational Measurement* is published by the Center for Educational Measurement. It is primarily concerned with educational testing in all of its many aspects. Distributed free of charge to Center members.

CONTENTS

Research Studies On CEM Tests

A. CET

BASILIO R. ILEDAN
OLIVIA G. ILAGAN

The primary purpose of standardized tests in any school setting is to help guidance counselors, teachers, and other school officers cope with placement, selection, classification, and diagnostic problems relating to the academic performance of students. The minimum requirement for any test to be useful in such decision-making activities is an empirical demonstration of its reliability and validity.

The validity referred to here is either predictive or concurrent validity. Technically, validity is expressed as a bivariate or multi-variate correlation coefficient of test scores and academic grades. The predictive validity of a test is investigated when it is used for selection purposes. Then the primary interest is the efficiency of the test in forecasting any measurable academic performance.

For a test to be predictive of performance in a particular school it should demonstrate the capacity to measure some abilities being rewarded in that institution, the reward being in terms of achievement grades. The higher the correlation between grades and test scores, the more confidence can be placed on the test as a valid predictor of academic performance.

Evidence of concurrent validity is sought when a test is proposed as a substitute for some information on the skills and knowledge that students are expected to possess at a given point in time. Thus, the magnitude of the correlation between test scores and the contemporary criterion chosen such as grades on a specific subject area (e.g., scores on the Chemistry Diagnostic Test and final grades in high school chemistry) is the measure of the concurrent validity of the test.

The College Entrance Test, or CET, was the first scholastic aptitude test developed at FAPE. The test was administered in 1971 to 45,000 students from private schools in 15 regional test centers all over the country, and to about 120,000 students in 1972 in 26 regional test centers. During this period less than 10 studies were undertaken which dealt with the predictive validity of the test in the different schools which used it. Most of the studies were done by graduate students who used data from institutions affiliated to the FAPE testing program. Correlations reported in the specific areas of English, Mathematics, and Science (Boncale, 1972; Goyena, 1972; Gumban and Iledan, 1972; Kerr, 1972; Laurico-Ignacio, 1972; Nazareno, 1972; Ravelo, 1972) ranged from .09 to .88. The correlations obtained between the CET overall scores and college averages ranged from .17 to .53, showing that on the average CET scores were better predictors of college performance than high school grades. Multiple correlations of high school grades, CET scores, and college grades ranged from .33 to .84.

The trend was that low validities were obtained on the CET in highly selective schools and high validities in average schools. This is partly a function of the fact that in item selection, item difficulty levels around p = 0.50 were used based on a general population of students. The samples came from the following schools: University of San Carlos, St. Paul College of Manila, Ateneo de Manila, St. Scholastica's College, Maryknoll College, St. Joseph College, College of the Holy Spirit, De La Salle University, University of Baguio, Columban College, Holy Angel College, Philippine School of Business Administration, Regina Carmeli, San Beda College, and Notre Dame of Jolo.

Plans for further research on the CET were set aside in 1973 due to the involvement of FAPE in the development of the NCEE, the nationwide testing program promulgated by Presidential Decree No. 6-A. During this period the groundwork for the development of other FAPE tests was laid out. But being the first product of FAPE's efforts at test development, there is a lingering sentiment over the CET. The question often asked about it is: "How valid was the CET over the years in the schools which elected to use it? " Only a major study will answer this question but which is not likely to be undertaken.

Out of curiosity and sentiment, however, a small sample of 129 students who took the CET in 1973 was obtained from the 1976-77 population of senior college students at the College of the Holy Spirit, an exclusive school for girls. Each student in the sample had three sets of information: fourth year high school average, CET scores, and three years of college grades. The high school average (HSA) and CET overall score (GSA) were correlated with college grades (GPA) by course and by level from first to third year. The results are presented in Table 1.

Table 1 GPA, HSA and GSA Intercorrelations by Level and by Degree Program, College of the Holy Spirit

Degree Program	Sample Size	HSA x GSA	GPA x HSA			GPA x GSA		
			Yr 1	Yr 2	Yr 3	Yr 1	Yr 2	Yr 3
Med. Tech.	30	.24	.45	.27	.24	.66	.51	.35
AB-BS Com.	58	.21	.19	.23	.10	.45	.38	.42
Fine Arts	41	.02	.28	.12	.17	.58	.48	.19

Table 1 shows low correlations between high school averages and CET overall scores (HSA x GSA). A comparison of the correlations between college averages and high school averages (GPA x HSA) on the one hand, and those between college averages and CET overall scores (GPA x GSA) on the other indicates the CET to be a better predictor of academic performance than high school grades. Note that the correlations between college averages and CET overall scores decreased across the years except in AB-BS Commerce where there is an increase from .38 in the second year to .42 in the third year. The differences in magnitude of the correlations were found not to be significant between first and second year but are significant ($\alpha = .05$) between the first and third year. This decrease in correlations may be explained as the result of the increasing concentration of the students in their major fields as they progress from year to year. This means that the CET was useful for predicting college performance at most during the first two years of college because of related content in both predictor and criterion, but no longer useful in the higher years for some degree programs. This is best illustrated in the Fine Arts course in which the correlation of .58 in the first year decreased to .19 in the third year. From subjects like English, Mathematics, History, and some art and design fundamentals in the first year, students in the course progress to subjects like clay modeling, sculpture, poster and textile design, and cartooning in the upper years. Also, no significant difference was found between correlations across courses in the first and second years. This means that in this particular school the CET was equally predictive of the academic performance of the general population up to the second year. Table 2 below shows the multiple correlations obtained when both high school grades and CET scores were combined as predictors of college performance.

Table 2 Combined Predictive Validity of High School Grades and CET Scores in the 1st and 2nd Years of College

Degree Program	GPA x HSA x GSA	
	First Year	Second Year
Medical Technology	.73	.53
AB-BS Commerce	.46	.41
Fine Arts	.64	.49

The CET item file became the source of the test items which went into the tests developed for the 1973 and 1974 NCEE. Data gathered on approximately 300,000 college-bound students who took the first NCEE was later written up. These included studies on demographic variables and institutional profiles done at the Guidance and Testing Division and the Educational Data Bank of FAPE, respectively. The latest study done by the CEM staff consists of five volumes entitled, "Profiles of the 1973 NCEE Examinees."

After the NCEE, FAPE undertook the development of the following tests: Philippine Aptitude Classification Test (PACT), College Scholarship Qualifying (CSQT), College Scholastic Aptitude Test (CSAT), and Diagnostic Tests (DT). These tests have been in the field for the last three years and, therefore, a lot of data can already be gathered on them. The problem of course is the validity of the criterion measures that will be used in validating these tests. Results of some preliminary studies are discussed in the succeeding articles.

REFERENCES

Boncale, Romeo P. SVD. "An Institutional Study of the FAPE-CET in the University of San Carlos: One Year Follow-Up, 1971-1972," **FAPE Review**, IV(1). Makati: Office of Reports, FAPE, 1973, pp 58-61.

Laurico-Ignacio, Razminda S. "Predictive Validity and Placement Significance of the FAPE-CET in the University of Baguio, 1971-1972," **FAPE Review**, IV(1). Makati: Office of Reports, FAPE, 1973, pp 44-50.

Goyena, Sr. Chantal de Joseph, SPC. "A Concurrent and Predictive Validity Study of the CET in St. Paul's College of Manila," **FAPE Review**, IV(1). Makati: Office of Reports, FAPE, 1973, pp 51-53.

Kerr, Rebecca C. "The Validity of the CET in Predicting the Academic Performance of the DLSC Freshmen in the First Year of School Year 1971-72," FAPE Review, IV(1). Makati: Office of Reports, FAPE, 1973, pp 54-57.

Gumban, Remigio S. and Iledan, Basilio R. "Some Predictive Validities of the CET Based on a Relatively Homogeneous Group of Schools." Research Bulletin, July 1971. FAPE-GTD, Makati.

Nazareno, Remedios P. "A Factor Analytic Study of the FAPE College Entrance Test in Relation to Other Intellective Factors." Unpublished Thesis. Ateneo de Manila University, 1972, 92pp.

Ravelo, Antonio D. "An Institutional Validity and Utility Study of the College Entrance Test in a Provincial College." Unpublished Thesis. University of Sto. Tomas, April 1972, 181pp.

B. CSAT

AURORA P. MINA
OLIVIA G. ILAGAN

This initial CSAT validity study is limited to the Adamson University, a member of the Center for Educational Measurement (CEM). The study investigated the validity of the CSAT in predicting the academic performance of freshmen in three courses, namely: Engineering, Architecture, and Accounting. Specifically, it tried to answer the following questions:

1. How valid is the CSAT in predicting college freshman performance at the Adamson University?
2. How predictive is the CSAT compared to the NCEE?
3. How valid is the CSAT in predicting performance in the different college subject areas?

The above questions are answered by testing the following hypotheses:

1. The correlation between college grade point average (GPA) and CSAT composite score (GSA) would equal or even exceed .30, the minimum value considered adequate.
2. The CSAT GSA is a better predictor of academic performance than the NCEE GSA.
3. There are significant differences between CSAT and NCEE component scores in predicting performance in corresponding college subject areas.
4. Multiple correlation indices obtainable in predicting college freshman GPA from the combination of CSAT GSA and NCEE GSA will be higher than the correlations obtained using either of these predictors.

The sample was limited to the freshmen of the Adamson University who took the CSAT in 1976-1977 as required

Table 1. Number and Percentage of "Incomplete," "Dropped," and "Not Taken" for Each Subject Area in the Freshman Year at the College of Engineering, Adamson University (N=1033)

Subjects	Incomplete		Dropped		Not Taken		Total	
	N	%	N	%	N	%	N	%
English 1	34	3.3	6	0.6	52	5.0	92	8.9
Humanities 1	26	2.5	11	1.4	53	5.1	90	8.7
College Algebra 1	53	5.1	28	2.7	40	3.9	121	11.7
Plane Trigonometry	47	4.6	21	2.0	39	3.8	107	10.4
Chemistry Lec 1	47	4.6	67	6.5	29	2.8	143	13.8
Chemistry Lab 1	47	4.6	19	1.8	46	4.4	112	10.8
Engineering Drawing 1	122	11.8	9	0.9	79	7.6	210	20.3
English 2	54	5.2	25	2.4	131	12.7	210	20.3
Humanities 2	29	2.8	6	0.6	186	18.0	221	21.4
Anal. Geometry	22	2.1	53	5.1	467	45.2	542	52.5
Solid Geometry & Spher. Trigo.	61	5.9	63	6.1	384	37.2	508	49.2
Chemistry Lec 2	6	0.6	146	14.1	496	48.0	648	62.7
Chemistry Lab 2	24	2.3	34	3.3	538	52.1	596	57.8
Eng'g Drawing 2	115	15.0	18	1.7	258	25.0	431	41.7

by the university. To control bias in college performance due to unequal academic loads, only those with complete subject loads in the first and second semesters were chosen. Thus, from an original sample of 2,841 students only 174 were included in the final sample as a result of the elimination of those with incomplete grades, dropouts, and those with less than the regular load of subjects. Data on these three categories using the Engineering subsample is given in Table 1.

Architecture and Accounting followed the same trend in the number of incompletes and dropouts for the first semester subjects.

The sample of 174 freshmen was used in the analysis. The breakdown of the sample to the different courses is as follows:

Engineering	115 or	66.1%
Architecture	32 or	18.4%
Accounting	27 or	15.5%
	174	100.0%

A further distribution of the Engineering subsample to the different fields is shown below:

Chemical Engineering	49 or	42.6%
Industrial Engineering	43 or	37.4%
Mechanical Engineering	20 or	17.4%
Electrical Engineering	2 or	1.7%
Civil Engineering	1 or	0.9%
	115	100.0%

A total of 15 criterion and predictor variables were included in the analysis. High school grades which are commonly used as predictors of college performance were not available. The predictors used in the analysis are the CSAT and NCEE scores; the criterion variables are college grades.

The criterion variables are as follows:

(1) College Average in English: COLENG
(2) College Average in Mathematics: COLMATH
(3) College Average in Science: COLSCI
(4) First Year College Average: COLGPA

The predictor variables are as follows:

(1) NCEE Reasoning Ability NCEE RA
(2) NCEE Mathematical Ability: NCEE MA
(3) NCEE Verbal Ability: NCEE VA
(4) NCEE Reading Comprehension: NCEE RC
(5) NCEE General Scholastic Aptitude: NCEE GSA
(6) CSAT English: CSAT E
(7) CSAT Mathematics: CSAT M
(8) CSAT Science: CSAT S
(9) CSAT Verbal Relations: CSAT VR
(10) CSAT Inductive Reasoning: CSAT IR
(11) CSAT General Scholastic Aptitude: CSAT GSA

The statistical analysis employed was stepwise multiple regression. Some of the indices yielded are Pearson product-moment correlations, beta weights, and multiple correlations. Intercorrelations of the predictors are given in Table 2. The other regression indices are given in Tables 3, 4 and 5.

Table 2. Intercorrelations of CSAT and NCEE Component Scores; Correlation Between CSAT and NCEE GSA's

CSAT	NCEE RA	NCEE MA	NCEE VA	NCEE RC
E	.51	.45	.64	.53
M	.46	.61	.38	.34
S	.45	.49	.50	.43
VR	.51	.29	.42	.43
IR	.38	.46	.34	.26
NCEE GSA x CSAT GSA: r = .67				

Table 3. Coefficients Obtained in the Prediction of First Year College Performance Using NCEE and CSAT Scores as Predictors: AU College of Engineering (N = 115)

Criterion	Predictor	Simple r	Multiple R	Beta Wt.
COLGPA	NCEE GSA	.6478	.6732	.480**
	CSAT GSA	.5718		.248**
COLENG	CSAT E	.5664	.6473	.371**
	NCEE RC	.5656		.369**
COLMATH	CSAT M	.4290	.4678	.300**
	NCEE MA	.3979		.227*
COLSCI	CSAT S	.3354		.335**

*significant at .05
**significant at .01

Table 4. Coefficients Obtained in the Prediction of First Year First Semester College Performance at the AU School of Architecture (N = 32)

Criterion	Predictor	Simple r	Multiple R	Beta Wt.
COLGPA	CSAT GSA	.70	.73	.817**
	NCEE GSA	.40		–.149
COLENG	CSAT VR	.45	.49	.264
	NCEE VA	.43		.188
	CSAT E	.39		.110
COLMATH	CSAT M	.69	.75	.446
	NCEE MA	.66		.384

Table 5. Coefficients Obtained in the Prediction of First Year First Semester College Performance of Accounting Students, AU College of Commerce (N = 27)

Criterion	Predictor	Simple r	Multiple R	Beta Wt.
COLGPA	NCEE GSA	.71	.76	.519**
	CSAT GSA	.62		.329
COLENG	NCEE VA	.66	.70	.412
	CSAT E	.65		.343
COLMATH	CSAT M	.70	.70	.744**
	NCEE MA	.38		–.073

The results of the hypotheses testing procedures are summarily presented below:

1. Engineering: The CSAT GSA when correlated with the first year college GPA yielded a correlation coefficient of .57 which suggests a positive and highly significant relationship between predictor and criterion. This exceeds the .30 value considered adequate for prediction. The same finding is true for the NCEE GSA which yielded a correlation of .65 with the college GPA. Both correlations are significant beyond the .01 level.
   Architecture: The CSAT GSA correlated .70 with the COLGPA while NCEE GSA correlated .40 with the same criterion. NCEE GSA seems to act as a suppressor in this case because the analysis yielded a negative beta weight of −.149.
   Accounting: Both predictors correlated highly with the criterion: .71 for NCEE GSA and .62 for CSAT GSA.
2. Engineering: Although NCEE GSA correlated higher with COLGPA than CSAT GSA the test showed that there is no significant difference between the values .65 and .57. Architecture: The reverse was observed for Architecture. The difference between the correlations .70 and .40 is significant at the .05 level.
   Accounting: There is no significant difference between the values .71 and .62.
3. Engineering: Highly significant relationships were found to exist between the CSAT component scores and college subject area averages. The simple r values ranged from .43 to .57. The NCEE component scores also followed the same trend. However, only moderate correlation was obtained when NCEE MA was correlated with college mathematics average grades.
   Architecture: High coefficients were obtained when CSAT component scores were correlated with college subject area averages in comparison with those obtained when NCEE component scores were correlated with the same criterion.
   Accounting: Correlations were high between CSAT component scores and corresponding college subject area averages, with r values of .65 for English and .70 for Mathematics. Only NCEE VA is predictive of college English with an r value of .66. The NCEE MA acted as a suppressor variable when regressed with college Mathematics and made no significant contribution to the multiple R (refer to Table 5).
4. Multiple R's of .73, .76, and .67 for Architecture, Accounting, and Engineering, respectively, were obtained in predicting college freshman GPA from CSAT GSA and NCEE GSA. When the two predictors are used instead of either one, the increase in the multiple R is significant at the .01 level. Higher multiple correlations were also obtained in predicting college subject area averages from corresponding CSAT and NCEE component scores except for Accounting where the addition of the NCEE MA scores did not improve the multiple correlation.

From the results of this study, the following tentative conclusions were drawn:

1. Both CSAT and NCEE are valid for prediction of first year academic performance in the three courses.
2. The CSAT component tests are valid for the prediction of performance in college English, Mathematics, and Science.
3. The combined use of the CSAT and the NCEE is a more reliable and precise procedure for admissions and placements than using only either one.

C. MULTICOLLINEARITY IN THE PREDICTOR VARIABLES

BASILIO R. ILEDAN
AURORA P. MINA

Problems arise in multiple regression when all or some of the independent variables or predictors are highly correlated (.50 and above). The greater the correlation between two predictors, the more unreliable the relative importance of the regression coefficients become, their values varying greatly from sample to sample (Kim and Kohout, 1975). This problem brought about by multicollinearity in the predictors is clearly illustrated in the case of the CSAT and NCEE scores when used together in the prediction of academic performance at the Adamson University, particularly in the area of English.

The correlations of the component scores of CSAT and NCEE with the average grades in freshman English at the College of Engineering of the Adamson University are given in Table 1. CSAT E has the highest correlation with the criterion COLENG, followed by NCEE RC. There is, however, no significant difference in their correlations with the criterion. The intercorrelations of the four predictors are given in Table 2. CSAT E is highly correlated with CSAT VR (.66) and NCEE VA (.64). NCEE RC and NCEE VA are also correlated (.61).

Table 1. Correlations of CSAT E, CSAT VR, NCEE RC, and NCEE VA with Average Grades in Freshman English (N = 106)

Predictor	Criterion: COLENG
CSAT E	.5664
NCEE RC	.5656
CSAT VR	.5310
NCEE VA	.4725

Table 2. Intercorrelations of CSAT E, CSAT VR, NCEE RC, and NCEE VA

Predictors	CSAT E	NCEE RC	CSAT VR	NCEE VA
CSAT E	–	.5292	.6578	.6440
NCEE RC		–	.4192	.6067
CSAT VR			–	.4270
NCEE VA				–

In multiple regression using the stepwise solution, the criterion for determining the relative importance of a predictor at the first step is the magnitude of its correlation with the dependent variable. Thus, for the predictors given in Table 2, CSAT E will be considered first in the solution as having the highest contribution in the prediction, and NCEE VA the last. A simplified outline of the stepwise multiple regression process involving the aforementioned variables is presented below:

Criterion: Freshman English Average (COLENG)

Predictor (s): NCEE VA, NCEE RC, CSAT E, CSAT VA

Step	Predictor Entered	Beta Wt.	Multi R
1. . .	CSAT E	.5664	.5664
2. . .	CSAT E NCEE RC	.3710 .3693	.6473
3. . .	CSAT E NCEE RC CSAT VR	.2285 .3458 .2355	.6709
4. . .	CSAT E NCEE RC CSAT VR NCEE VA	.2139 .3343 .2368 .0309	.6712

Note that in Step 3 of the outline the entry of CSAT VR has reduced the beta weight of CSAT E more than it did NCEE RC. As shown in Table 2 the highest correlation is between CSAT E and CSAT VR, higher than the .57 correlation between CSAT E and the criterion COLENG. The correlation between NCEE RC and CSAT VR is only .42. The beta weight of CSAT E was reduced from .3710 in Step 2 to .2285 in Step 3, which is less than the beta weight of .3458 of NCEE RC. Normally, this does not happen if the predictors have low intercorrelations. The predictor having the highest correlation with the criterion usually has the highest beta weight (beta weights are reduced when another predictor is added to the set) regardless of the subsequent entry of other predictors.

Although it would seem desirable to include CSAT VR as one of the predictors of performance in English since its beta weight of .2355 in Step 3 is even greater than that of CSAT E, its contribution to the explained variance is actually not significant. The multiple correlation R of .65 in Step 2 increased to only .67 in Step 3. The obtained F ratio from these values is 2.87 which is not significant at the .05 level (df: 2 and 102). The same is true for NCEE VA when this is considered for inclusion as one of the predictors. The only variables, therefore, which can be considered as predictors of freshman English are CSAT E and NCEE RC. The increase of the multiple R from .57 in Step 1 to .65 in Step 2 yields an F ratio of 17.41 which is significant beyond the .01 level (df: 1 and 103). Also, it seems that it is only in Step 2 where the beta weights can be considered reliable. It is possible that at this point the existing multicollinearity has not yet affected the weights due to the fact that the correlation of CSAT E and NCEE RC is less than the correlation of either one with the criterion.

The unreliability of regression coefficients because of multi-collinearity in the predictors may be avoided by either combining the highly correlated variables, or using only one of them (Kim and Kohout, 1975). The scores on the two tests may be added before using them in the regression if their correlation is high enough to warrant it. This will simplify the process of prediction. This suggestion, however, needs empirical evidence before it is implemented.

REFERENCE

Kim, J. and Kohout, F.J. Multiple Regression Analysis: Subprogram Regression. **Statistical Package for the Social Sicences.** New York: McGraw-Hill, 1975, p. 340.

D. CONCURRENT VALIDITIES OF THE CEM DIAGNOSTIC TESTS IN SEVEN SCHOOLS

ELIZABETH S. SAMIA
ESTELITA B. ESPIRITU

The Diagnostic Tests (DT) are designed primarily to pinpoint weak points in learning outcomes for which remedial measures are needed. The tests consist of subtests which measure various content and behavioral dimensions such that a student's performance profile on a test will show his relative strengths and weaknesses on the corresponding academic subject. Thus, differential instructions can be resorted to in order to help the student achieve the most during the learning process. However, when a test is used for such decision-making purposes its content and concurrent validities should be well established.

For the Diagnostic Tests, content validity would refer to the evidence of agreement between the contents of the tests and the contents and curriculum objectives of the subjects being tested. Content validity is thus best insured by entrusting test construction to persons well-qualified to judge the relationship of test content to curriculum objectives. For each of the Diagnostic Tests, CEM engaged the services of a committee of subject area specialists who were knowledgeable in the various textbooks, curricula, syllabi, practices, opinions, and philosophies in each area and level of secondary education. These committees worked closely with the CEM staff in selecting and defining the topics and behavioral objectives that would be covered in the tests, and reviewing and revising when necessary the items written for the tests. The items were written by highly qualified and competent high school and college instructors. Each item was reviewed by the committee members assigned to the particular subject area (including the committee chairman as well as the CEM staff) before it was pretested on a stratified random sample of third year high school students to determine its statistical properties. In this way, every possible care was taken to make the Diagnostic Tests reflect the appropriate content in each subject area as implemented by Philippine high schools following the Revised Secondary Curriculum.

Institutional validity studies using data gathered from seven schools in the Greater Manila Area have been performed to establish the concurrent validity of the Diagnostic Tests. These schools are the Immaculate Conception Academy in Manila, the Immaculate Conception Academy in San Juan, Malate Catholic School, Notre Dame of Manila, St. Jude Catholic School, St. Paul's College in Parañaque, and St. Scholastica's College. Overall scores in each of the Diagnostic Tests and the final academic grades in Communication Arts (English), Social Studies, Science, and Mathematics of the third year students in these schools were collected. In addition, a listing of textbooks being used by these institutions and a copy of their course guides or syllabi were obtained. These information were very useful in establishing the correspondence between curricular and test content for these schools.

Systems of grading used by the schools were also noted. Cumulative and averaging systems of grading could affect the correlations obtained between the grades and the Diagnostic Tests scores. The Diagnostic Tests are administered at the end of the school year to obtain a measure of performance on the curricular contents of the core subjects taken up during the whole year. Thus, it is expected that the scores should have a high correlation with cumulative grades which take into account not only the performance of the students during the grading period when it is given but also previous grading periods. Therefore, the final grade is an indicator of the students accumulated progress. All except three of the schools sampled (Immaculate Conception Academy in Manila and in San Juan, and Malate Catholic School) use the cumulative grading system. The rest use the averaging system. In averaging, the grade for every period is independent of any grade given before. The final grade is the average of all these independent periodic grades.

Correlational analysis was used to establish the relationship between the Diagnostic Test overall scores and the final grades in corresponding subjects. The results of the analysis are shown in Table 1. The English Diagnostic Test demonstrated relatively high correlations ranging from .56 to .71 across the seven schools. For the Geometry Diagnostic Test, low correlations of .10 and .17 were obtained for St. Paul's College of Parañaque and Malate Catholic School while the other five schools have relatively high correlations ranging from .58 to .74. The Social Studies Diagnostic Test, on the other hand, had relatively high validity coefficients ranging from .39 to .68 in five schools, while low correlations were obtained in two schools, namely: Immaculate Conception Academy of Manila (.22) and St. Scholastica's College (.23). For Chemistry Diagnostic Test, a low correlation was obtained in St. Scholastica's College (.22) while correlations in the other six schools ranged from .37 to .70.

In interpreting the validity coefficients, caution should be used because the grades were those given at various high schools with different grading systems, curricula, teaching methods, and textbooks. However, it appears that the grading system may not be the main contributing factor to an obtained low correlation coefficient. This can be construed from the observed trend that there is no uniform grading system in the schools which have low correlations in the different subject areas.

The low correlation in Chemistry (.22) obtained in St. Scholastica's College may be attributed to the fact that the school offers an Integrated Science course in the third year where Physics and Biology are taught with Chemistry. Thus, the final grade obtained in this subject does not represent performance in Chemistry alone but of the other science subjects as well. In this case, there is a discrepancy between the criterion and the overall score in the Chemistry Diagnostic Test. The absence of a distinct final grade in Chemistry puts into question the adequacy of the criterion used.

Differences in textbooks being used cannot be assumed to have significant effects on the correlations obtained since the schools have a common list of books being used as textbooks and references.

Therefore, the differences in the concurrent validities of the Diagnostic Tests across the seven schools lie on the combined effect of the differences in curricula, grading systems, and teaching methods used.

Table 1. Validity Coefficients Obtained Between Diagnostic Tests Overall Scores and Final Subject Grades

Diagnostic Test	School	Sex of Sample	Sample Size	Validity Coefficient
English	Immaculate Conception Academy, Manila	F	108	.57
	Notre Dame of Manila	M	116	.64
	St. Jude Catholic School	M & F	149	.60
	St. Scholastica's College	F	176	.57
	Malate Catholic School	M & F	415	.56
	St. Paul's College, Parañaque	F	265	.61
	Immaculate Conception Academy, San Juan	F	116	.71
Geometry	Immaculate Conception Academy, Manila	F	108	.59
	Notre Dame of Manila	M	116	.74
	St. Jude Catholic School	M & F	151	.66
	St. Scholastica's College	F	177	.65
	Malate Catholic School	M & F	411	.17
	St. Paul's College, Parañaque	F	256	.10
	Immaculate Conception Academy, San Juan	F	120	.58

Diagnostic Test	School	Sex of Sample	Sample Size	Validity Coefficient
Social Studies	Immaculate Conception Academy, Manila	F	108	.22
	Notre Dame of Manila	M	113	.68
	St. Jude Catholic School	M & F	149	.60
	St. Scholastica's College	F	169	.23
	Malate Catholic School	M & F	411	.63
	St. Paul's College, Parañaque	F	266	.51
	Immaculate Concepcion Academy, San Juan	F	120	.39
Chemistry	Immaculate Conception Academy, Manila	F	108	.37
	Notre Dame of Manila	M	114	.59
	St. Jude Catholic School	M & F	146	.70
	St. Scholastica's College	F	170	.22
	Malate Catholic School	M & F	415	.52
	St. Paul's College, Parañaque	F	264	.48
	Immaculate Conception Academy, San Juan	F	118	.53

The Factor Structure of The Revised College Scholarship Qualifying Test

BASILIO R. ILEDAN
AURORA P. MINA

In the first factor analytic study of the College Scholarship Qualifying Test, or CSQT, four factors were identified, namely: Verbal Comprehension, Science Concepts, General Reasoning, and Spatial Relations (Iledan, 1976). The first two factors were interpreted as educational achievement trait factors and the latter two as primary aptitude factors. Verbal Comprehension, which was found to be the dominant factor, relates to the knowledge and application of the English language which the Filipino student has to familiarize himself with since it is the medium of instruction in his academic setting. The Science Concept factor represents a knowledge trait in the sciences at the secondary level of education. The two aptitude factors, however, were not clearly interpreted because of the confounding effects of the other tests which loaded on the two factors, tests which normally are not associated with the said factors.

These findings, therefore, indicated a need to improve the test, specifically, the reduction of its verbal content in order to strengthen the other factorial components such that clear-cut interpretations of the factors are achieved. This will entail a revision and restructuring of the contents of the test.

Test Revision

The revision of the CSQT involved the removal of some item types from the battery and the addition of new ones without necessarily altering the basic content specifications of the test. The contents of the first form of the CSQT which may be considered the prototype, and the revised form are given in Table 1. The CSQT prototype form was administered to 1141 COCOFED scholarship applicants in March, 1976. The revised form was administered to 1751 COCOFED scholarship applicants in March, 1977.

In the revised form the English test was replaced with Verbal Aptitude. This consists of the Reading Comprehension items from the English test, and the Verbal Analogies and Logical Reasoning (renamed Conclusions) items from the mental ability tests of the prototype form. The removal of the major portion of the English test was the strategy employed to reduce the verbal component of the CSQT without violating the assumptions in the proposed content specifications for the prototype form of the test (Parocha, 1976). Some of the items were replaced to improve the distribution of the difficulty indices but the content of the replacements were the same as the ones replaced. Testing time for both forms are the same – 2 hours and 45 minutes.

The Mathematics test was regrouped into two types of items, verbal and symbolic problems. The Science test was also regrouped into Science Information and Data Interpretation items. This regrouping was based on the intercorrelations of the subscores in the first study.

The two new figural tests added to the mental ability part are the Assembly and Patterns tests which have been independently pretested some years before in some of the test development activities of the FAPE. The revised mental ability

Table 1. Nomenclature of the Different Subtests and Item Types in the Prototype and Revised Forms of the CSQT

Prototype	Revised
Booklet I	Booklet I
A. English Test	A. Verbal Aptitude Test
1. Vocabulary	1. Verbal Analogies
a. synonyms	2. Conclusions
b. antonyms	3. Reading Comprehension
2. Grammar and Usage	B. Mathematics Test
3. Reading Comprehension	1. Verbal Problems
B. Mathematics Test	2. Symbolic Problems
1. Geometry	C. Science Test
2. Arithmetic	1. Science Information
3. Algebra	2. Data Interpretation
C. Science Test	
1. General Science	Booklet II
2. Biology	A. Figural Reasoning
3. Chemistry	B. Assembly
4. Physics	C. Number Series
Booklet II	D. Patterns
A. Figural Reasoning	E. Hidden Figures
B. Logical Reasoning	
C. Number Series	
D. Verbal Analogies	
E. Hidden Figures	

part of the CSQT as seen in Table 1 consists of four figural tests and one symbolic test.

Data Analysis

Twelve raw scores were obtained on the revised CSQT based on the item classification given in Table 1. The interest of the authors is in finding explanatory concepts for the observed relationships among these twelve scores. Note that the values in the correlation matrix in Table 2 are more or less equal in magnitudes which makes it difficult to interpret the relationships among the variables based on their inter-correlations. Factor analysis was therefore done to discover the factor structure in the revised CSQT which may be different from the factor structure obtained in the first study.

The particular method of data reduction employed is the principal axis solution followed by a varimax rotation to obtain a simple structure in the factor matrix, and yield factors that are at the same time orthogonal. Principal axis gives the best linear summary of the variance of a test score explained by the set of factors obtained in the analysis. The varimax method of rotation (Kaiser, 1956) maximizes the factor loadings of each variable in the least number of factors within the factor matrix, ideally in only one if possible, in order to achieve the closest approximation to simple structure.

Using the first study as the basis, five factors were extracted from the correlation matrix. The scree test (Cattell, 1966) was applied to the eigenvalues of the principal axis factor matrix shown in Table 3.

Table 2. Intercorrelations of the 12 CSQT Subscores*

Mnemonics	1	2	3	4	5	6	7	8	9	10	11	12
VA	–	41	43	55	47	53	45	43	38	38	30	31
CON		–	34	41	38	42	34	31	28	29	20	22
RC			–	40	34	45	33	26	24	26	20	24
VM				–	68	59	51	48	45	54	33	37
SM					–	53	47	46	44	52	33	37
SI						–	61	38	39	36	32	36
DI								38	40	36	32	36
FR								–	48	45	34	36
AS									–	36	38	35
NS										–	26	32
PAT											–	38
HF												–

*Decimal points omitted

Table 3. Scree Test Indicating Five Salient Factors

Factor	% Total Variance	Eigenvalue	Difference Between Eigenvalues
I	44.4	5.53	
			4.25 (slope)
II	9.0	1.08	
			0.23 (slope)
III	7.1	0.85	
			0.16 (slope)
IV	5.8	0.69	
			0.01 (slope)
V	5.7	0.68	
			0.04*
VI	5.3	0.64	
			0.03
VII	5.1	0.61	
			0.08
VIII	4.4	0.53	scree
			0.04
IX	4.1	0.49	
			0.04
X	3.8	0.45	
			0.10
XI	2.9	0.35	
			0.05
XII	2.5	0.30	0.0

*First reversal point

This test considers as meaningful all factors before the first point of reversal in the decreasing differences of consecutive eigenvalues.

Another criterion for determining the significance of factors is based on the size of their eigenvalues. Only factors with eigen values of 1.0 or greater are considered significant since 1.0 is the amount representing the total variance of a single variable. If this criterion is used then only two orthogonal factors can be considered significant in the present analysis as noted from the eigenvalues in Table 3.

However, the purpose of the analysis as already mentioned is to discover the underlying factor pattern inherent among the tests which compose the CSQT. The importance of this is in considering the individual merits of each test with respect to the ability that each measures. The number of factors indicates as salient in the scree test was therefore considered.

Two analyses were done. The first exploratory factoring indicated that the Number Series test loads substantially on the same factor as the verbal and symbolic Mathematics tests. This can be verified from the varimax rotated factor matrix presented in Table 4b. This result substantiates a similar finding in the first study (Iledan, 1976). Therefore, a second factoring was done with the Number Series test deleted from the matrix since it appears to duplicate the function of the Mathematics tests and does not correlate substantially with any of the other factors to warrant its retention in the battery.

The varimax rotated factor matrix of the second analysis is given in Table 5b. Although the five factors are very distinct, it must be pointed out that they do not substantially explain the variances of some of the variables as indicated by their obtained communalities, h^2. For instance, the obtained communalities of the Conclusions and Patterns subtests are the lowest, 32% and 33%, respectively, as seen from Table 5a. These are the maximum communalities obtainable for the two variables as far as the five factors are concerned. This means that more than 65% of what ever these two subtests measure are not within the abilities domain defined by the five factors, and may be attributed to unique variance. Table 6 gives the proportion of total variance accounted by the five factors.

Table 4a. Principal Axis Factor Matrix with the Number Series Test Included

Variable Name	Mnemonics	FACTOR I	II	III	IV	V	h^2
Verbal Analogies	VA	.68	–.15	.03	.22	–.02	.54
Conclusions	CON	.52	–.17	.00	.14	.00	.32
Reading Comprehension	RC	.51	–.26	.10	.21	.12	.40
Verbal Mathematics	VM	.80	–.09	–.25	–.08	.03	.72
Symbolic Mathematics	SM	.76	.03	–.25	.14	.12	.66
Science Information	SI	.75	–.33	.19	–.18	–.09	.76
Data Interpretation	DI	.67	–.08	.17	–.17	–.06	.52
Figural Reasoning	FR	.64	.26	–.08	.14	–.18	.54
Assembly	AS	.60	.25	.05	.00	–.18	.46
Number Series	NS	.61	.13	–.24	.01	–.07	.45
Patterns	PAT	.48	.25	.19	–.02	.00	.33
Hidden Figures	HF	.52	.32	.26	–.01	–.24	.50
Eigenvalue		4.88	.55	.37	.21	.17	6.20
Percent of Total and		40.71	4.58	3.08	1.75	1.42	
Common Variance Explained		78.71	8.87	5.97	3.39	2.74	

Table 4b. Varimax Rotated Factor Matrix with the Number Series Test Included

Variable Name	Mnemonics	FACTOR I	II	III	IV	V
Verbal Analogies	VA	.57*	.26	.18	.19	.28
Conclusions	CON	.45*	.23	.10	.17	.16
Reading Comprehension	RC	.57*	.15	.15	.15	.05
Verbal Mathematics	VM	.39	.64*	.14	.30	.23
Symbolic Mathematics	NVM	.28	.66*	.24	.25	.19
Science Information	SI	.49	.25	.17	.64*	.14
Data Interpretation	DI	.31	.25	.29	.48*	.20
Figural Reasoning	FR	.23	.34	.27	.09	.54*
Assembly	AS	.16	.25	.32	.22	.47*
Number Series	NS	.22	.53*	.21	.07	.25
Patterns	PAT	.13	.14	.43*	.17	.28
Hidden Figures	HF	.17	.18	.64*	.10	.14

*Highest factor loading for the variable.

Table 5a. Principal Axis Factor Matrix with the Number Series Test Deleted

Variable Name	Mnemonics	FACTOR I	II	III	IV	V	h^2
Verbal Analogies	VA	.69	–.14	.05	.21	–.02	.54
Conclusions	CON	.52	–.16	.02	.13	.00	.32
Reading Comprehension	RC	.52	–.24	.15	.19	.11	.40
Verbal Mathematics	VM	.79	–.12	–.25	–.04	.03	.71
Symbolic Mathematics	SM	.75	–.00	–.31	–.11	.16	.70
Science Information	SI	.76	–.27	.17	–.18	–.08	.72
Data Interpretation	DI	.68	–.05	.16	–.21	.08	.55
Figural Reasoning	FR	.63	.23	–.10	.14	–.14	.49
Assembly	AS	.61	.28	–.04	.03	–.18	.48
Patterns	PAT	.49	.29	.11	–.01	.01	.33
Hidden Figures	HF	.52	.35	.21	–.02	.23	.49
Eigenvalue		4.52	.53	.31	.21	.16	5.73
Percent of Total and Common		41.09	4.82	2.82	1.91	1.45	
Variance Explained		78.88	9.25	5.41	3.66	.79	

Table 5b. Varimax Rotated Factor Matrix with the Number Series Test Deleted

Variable Name	Mnemonics	FACTOR I	II	II	II	V
Verbal Analogies	VA	.58**	.21	.17	.31	.18
Conclusions	CON	.46*	.20	.10	.19	.17
Reading Comprehension	RC	.57*	.13	.16	.06	.15
Verbal Mathematics	VM	.42	.58*	.15	.31	.27
Symbolic Mathematics	NVM	.30	.66*	.25	.26	.20
Science Information	SI	.49	.26	.17	.17	.60*
Data Interpretation	DI	.31	.22	.28	.23	.52*
Figural Reasoning	FR	.25	.24	.27	.54*	.10
Assembly	AS	.16	.20	.30	.53*	.20
Patterns	PAT	.13	.12	.43*	.31	.16
Hidden Figures	HF	.17	.13	.63*	.17	.11

*Highest factor loading for the variable.

Table 6 Proportion of Variance Attributed to Each of the Varimax Rotated Factor

	Factor Name	Mnemonics	% Variance Extracted
I.	Verbal Reasoning in English	VRE	14.69
II.	General Mathematics	GM	10.16
III.	Spatial Perception 1	SP1	9.11
IV.	Spatial Perception 2	SP2	9.81
V.	Science Concepts	SC	8.32
		TOTAL	52.00

Interpretation of the Factors

The interpretation of the factors will be based on the factor loadings in the varimax matrix presented in Table 5b. Generally, the minimum factor loading considered noteworthy is .30 (Nunnally, 1967) which represents about 10% of the variance. The importance of a factor for any of the variables involved in the factoring is expressed by the amount of variance in the test accounted for by the factor. This variance is simply the square of the factor loading which represents the correlation between the variable and the factor. For example, the variance of Verbal Analogies (see Table 5b) accounted for by Factor I is $(.58)^2 = .3364$ or about 33.6%. Factor II accounts for 4.4%, Factor III for 2.9%, Factor IV for 3.6%, and Factor V for 3.2%. These all add up to 54% which is the obtained communality, h^2, of Verbal Analogies. Thus, 46% of the variance of Verbal Analogies cannot be accounted for by the five factors and is then considered to be unique variation.

The interpretation of each factor will be based on published findings, particularly on Lohnes's definition of ability and its hierarchical classification. In his outline of a theory in trait and factor psychology, Lohnes states:

"We define a **trait** as an enduring pattern of behaviors which is exhibited by many people, but in varying degrees. . measurable. . classified as either an **ability** or a **motive**..

"Ability is the generic term for a domain of traits which can be further classified as **general intelligence, knowledges, and aptitudes.** General intelligence is a very pervasive trait that facilitates quickness and quality of responses to all cognitive tasks to some degree. . . operates to condition almost all school learning. . .

"A knowledge is a performance trait that enables the subject to reproduce associations or to complete gestalts from a broad class of cognitive holdings. . . generate and apply information in subject-matter area. . depends more on specific learning opportunities. . .

"An aptitude is a performance trait that facilitates speed and precision of response to items from a specific, unique class of relatively simple tasks. . . ." (Lohnes, 1966)

In the above definitions, general intelligence is at the top of the hierarchy in the order of complexity. None of the five factors in the analysis qualifies as general intelligence since there are not enough indicators for this factor in the CSQT. Three of the factors represent knowledge or educational achievement traits, and the other two represent aptitude traits. These factors can only be understood in terms of the contents of the variables which loaded meaningfully in each of them.

Factor I: **Verbal Reasoning in English.** The most important of the five factors is the Verbal Reasoning in English, or VRE, which has the highest share of the total variance of 52% accounted for by the five factors. It accounts for 14.7% of the variance to which seven of the eleven subscores have meaningful contributions, i.e., they correlate .30 and higher with

VRE. The contents of these subtests identify VRE as a knowledge factor. A knowledge trait has been defined as the ability to generate and apply information about a subject-matter area.

The subtest which correlated highest with VRE is Verbal Analogies. One of the items in this subtest is:

intelligent : dull

(A) beauty : brains
(B) fluent : clear
(C) strong : weak
(D) jolly : happy

It is apparent that the ability to respond to this type of question depends on some knowledge and comprehension of English words and their implied relationships.

The second highest correlation with VRE is that of Reading Comprehension. This subtest consists of context-dependent items. The selections on which items are based are excerpts from literary pieces or articles on almost any subject matter, e.g., science, economics, humanities. The first item in one of the selection is:

The selection was probably written to make the reader

(A) aware of the limited power of the sun
(B) appreciate the importance of sunlight to life
(C) realize at least the beauty of the sun
(D) realize at least why man's ancestors worship the sun

The questions may be based on explicit statements in the selection, on explanations, generalizations, conclusions, or implications such as the one given above.

The third highest loading in the VRE factor is that of Conclusions. One item in this subtest is:

The government wants to succeed in creating new jobs for the country's unemployed. It will have to find a large amount of money for the purpose.

(A) The government is unable to solve the unemployment problem.
(B) Unemployment in the country is increasing every year.
(C) The government needs money in order to create new job opportunities.
(D) Unemployment is one cause of the "brain drain" in this country.

Questions in this subtest are based on everyday news and articles from newspapers and magazines. They are similar in construction to the items in the Inference Test of the Kit of Reference Tests for Cognitive Abilities (French et al, 1963) which is a marker test for the syllogistic or deductive reasoning factor.

It is clear that the type of items in the subtests which identify the VRE factor require the ability to see relations and implications in words and systems of words (sentences), the ability which Guilford and Hoepfner (1971) suggested are involved in deductive thinking.

Four other subtests which correlate substantially with VRE, although not their highest, are Science Information, Verbal Mathematics, Data Interpretation, and Symbolic Mathematics due to the verbal component in their items. These subtests deal on subject-matter areas communicated in the English language. The factor VRE, therefore, is defined as the ability to draw inferences or conclusions from the relationships and implications of given words or statements in the English language. It is an ability classified as knowledge and subsumes the comprehension of English as a learned language since it has been and still is the medium of instruction in the Philippine educational system.

Factor II: **General Mathematics**. The second most important factor is also an educational achievement trait factor located by the mathematics subtests, Verbal Mathematics and Symbolic Mathematics. It has been given the name General Mathematics and abbreviated to GM. None of the other subtests has a meaningful loading on this factor. The highest correlation is that of the Symbolic Mathematics subtest which consists mostly of algebraic and geometry problems largely stated in the usual symbolic language of mathematics. Two items in the subtest are:

1. $\frac{3x^2}{6x} =$

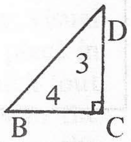
   (A) $2x$ (B) $\frac{x}{2}$ (C) $\frac{2x}{3}$

   (D) $\frac{1}{2x}$

2. In $\triangle$ BDC, BD =

   (A) 7 (C) 5

   (B) 25 (D) $5\sqrt{2}$



Three of the items in the Verbal Mathematics subtest are:

1. The algebraic translation of the statement "the square of a certain number is seven more than six times the number," is

   (A) $x^2 = 6x + 7$

   (B) $x^2 = 7 - 6x$

   (C) $x^2 + 7 = 6x$

   (D) $x^2 = 6x - 7$

2. If the measure of the vertex angle of an isosceles triangle is 100^o, then each of the base angle has a measure of

   (A) 100^o (C) 40^o

   (B) 80^o (D) 30^o

3. If four parts of rice are mixed with one part of corn, what part of the mixture is rice?

   (A) $\frac{1}{4}$ (B) $\frac{5}{4}$ (C) 4

   (D) $\frac{4}{5}$

The items in the Mathematics tests are aimed at measuring facility in the application of some mathematical operations. The contents of the tests are very similar to those of the mathematics tests in the Project TALENT* battery. The authors, therefore, are inclined to agree with Lohnes's (1966) interpretation of a similar factor. GM represents an educational achievement trait; it is the ability to organize and apply knowledge of basic mathematical concepts to the relevant aspects of problems in elementary mathematics and reasoning through to find solutions for them.

It must be pointed out that this factor was tentatively identified as general reasoning in the initial study (Iledan, 1976). This was due to the fact that some of the test items in the CSQT Mathematics tests are similar to those in the two Mathematics tests in the French kit which are marker tests for the general reasoning factor. However, the interpretation was not clear-cut because the Number Series test, a well-known marker test for the induction factor, had its highest loading on the general reasoning factor. Results from the initial study and the present one seem to verify the doubts expressed by Cattell (1971) and Pawlik (1966) that general reasoning and induction are not separate factors. Whether they are one or separate, the implication in the present study is that both are subsumed under the GM factor, i.e., the general reasoning and induction aptitudes must necessarily be components of the more complex factor GM.

Factor V: **Science Concepts.** The third factor in the study identified as a knowledge factor is Science Concepts, or SC. This is the same factor identified in the initial study of the CSQT. The SC factor is located by the Science Information and Data Interpretation subtests. One of the items in the former is:

What is the effect of stimulants in the body?

(A) speed up the body's activities
(B) slow down the body's activities
(C) produce psychological effects on the user
(D) cause increase in weight

---

*Project TALENT, a cooperative project between the University of Pittsburgh and the United States Office of Education, is a longitudinal study of the American youth which started in the late 1950's under the leadership of Dr. John C. Flanagan.

Most of the Data Interpretation items depend on tabulated data, and one such item is:

A class was studying the factors that affect the number of swings per minute of a pendulum. They varied the length of the string and kept the other factors the same. The following results were obtained:

length of string in inches	number of swings per minute
30.0	38
15.0	50
7.5	69

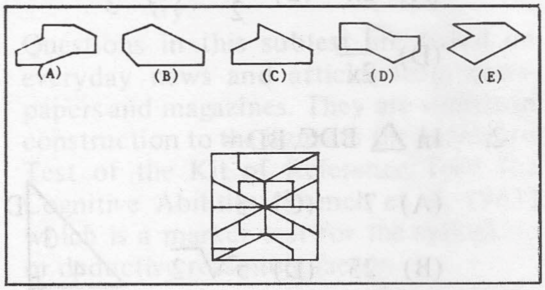
Which of the following generalizations apply to the results?

(A) The length of the string is inversely proportional to the number of swings per minute.
(B) The length of the string is directly proportional to the number of swings per minute.
(C) The length of the string does not affect the number of swings per minute.
(D) The number of swings is independent of the size of the string used.

No changes were made in the Science test except for the reclassification of its items into those which elicit knowledge and comprehension of learned information in the sciences at the secondary level of education, and those which provide data (either verbal, numerical, or pictoral) for interpretation or from which conclusions have to be drawn. SC is a rather narrow factor considering the fact that it is the last in order of importance in the matrix given in Table 5b. Also, because a substantial portion of the variances of the two subtests (24% and 10%) are already explained by the VRE factor.

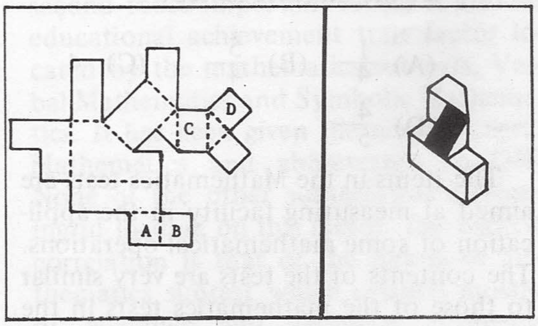
Factor III: **Spatial Perception 1**. The first of the two aptitude factors is Spatial Perception 1, or SP1. This factor is similar to the flexibility of closure factor defined by Royce (1973) as "the ability to 'hold in mind' a particular visual configuration and find it embedded in distracting material. Catell (1971) concludes that this factor is one of the primary factors that make up the second order factor visualization. SP1 is tapped by the Hidden Figures, Patterns, and Assembly subtests, with the first two having their highest loading on this factor.

One item in the Hidden Figures subtest is:



The task required of the subject is to choose which of the five simple geometric figures is embedded in the given figure made complex by criss-crossing lines. The Hidden Figures subtest was modeled after the Hidden Figures test in the French Kit. The test is an indicator of the flexibility of closure factor. It has also been variously interpreted as an indicator of fluid ability (Cattell, 1971), of convergent production of figural transformation (Guilford, 1967), and of the cognitive style factor field-independence (Witkins, 1976). Field independence refers to an individual's analytic modes of perception, an inherent characteristic of individuals who are naturally endowed with mental faculties capable of understanding more complex concepts in science, mathematics, and other related fields.

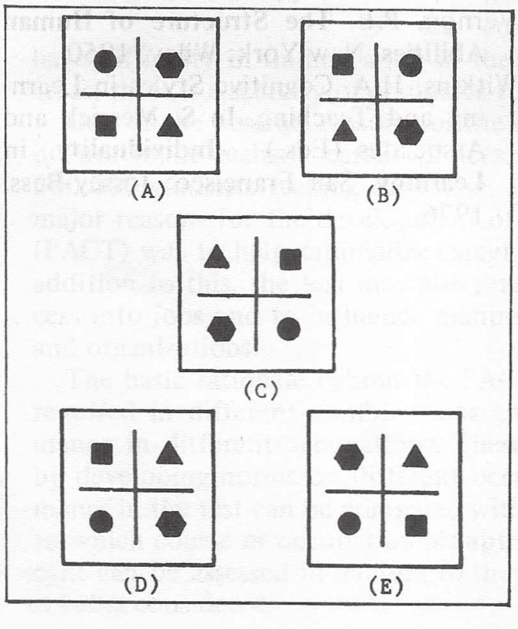
The Patterns subtest consists of three-dimensional geometric figures with corresponding two-dimensional patterns. One of the items in the subtest is:



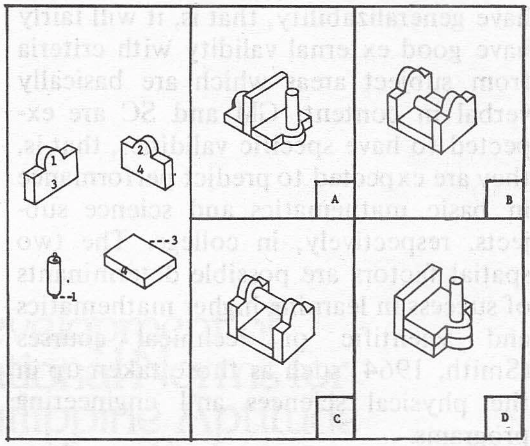
The task called for is to identify which of the four parts labeled A, B, C, and D in the pattern corresponds to the shaded part in the three-dimensional figure on the right. In a sense, the shaded part is also embedded in distracting material when this is being identified in the pattern.

Factor IV: **Spatial Perception** 2. The second aptitude trait factor has been named Spatial Perception 2, or SP2. It seems to be similar to the spatial orientation factor defined by Royce (1973) as "the ability to put together by visual imagination parts that are out of place in a visual pattern and to identify such 'out of place' precepts." Marker tests in the French Kit for spatial orientation are the Card Rotations and Cube Comparisons tests. The first requires the subject to detect whether or not cards of various configurations drawn on paper have been turned over as well as rotated. The second marker test requires the subject to determine whether or not cubes illustrated on paper are the same where the choices have been rotated to different positions.

Figural Reasoning and Assembly are the two subtests which have their highest loadings on SP2. These two subtests have similar elements as the two marker tests mentioned above and require the same visual manipulations. One of the items in the Figural Reasoning subtest is:



The task is to discover the principle which makes four of the five given figures similar to one another, and one of them the "out of place" figure, and consequently the right answer to the item. In the Assembly subtest the task is to choose the correct assembly from the four given assemblies of a set of dismantled parts of a geometric object. One of the items in the subtest is:



Three other subtests have meaningful loadings of .31 on the SP2 factor. These are Verbal Analogies, Verbal Mathematics, and Patterns. One might ask what the common denominator is between these three subtests and the two highest indicators of SP2. All of these subtests are potentially rich sources of visual imagery which is the reason why Verbal Analogies, which has been known to load on spatial factors loaded on this factor. For some of the items, the subject has to translate the verbal relationships into visual images to get at the right answer. Naturally, the same kind of ability would be required in a mathematics problem which verbally communicates ideas that may be essentially geometric or spatial. In the Patterns subtest the need for such an ability is more pronounced since the subject has to visually fold the two-dimensional pattern to form the three-dimensional figure before he can discover which of the labeled portions in the pattern corresponds to the shaded face in the figure.

Conclusions

This factor analysis has identified five relatively uncorrelated factors in the factor structure of the CSQT. Three of the factors represent knowledge or educational achievement traits; the other two represent spatial aptitude traits. The ultimate value of these factors, however, is in their predictive validity.

Of the five, only VRE is expected to have generalizability, that is, it will fairly have good external validity with criteria from subject areas which are basically verbal in content. GM and SC are expected to have specific validities, that is, they are expected to predict performance in basic mathematics and science subjects, respectively, in college. The two spatial factors are possible determinants of success in learning higher mathematics and scientific or technical courses (Smith, 1964) such as those taken up in the physical sciences and engineering programs.

The analysis also indicated that the test has only two significant factors if the criterion used for significance is an eigenvalue of 1.0 or greater. Perhaps this can be taken as a suggestion that the CSQT should have two composite scores, one representing general scholastic aptitude as measured by the VRE, GM, and SC factors, and the other representing visual reasoning as measured by SP1 and SP2. According to Vernon (1950), "the perception of form is a general characteristic of the abstract thinking involved in mathematics and science. . . ." These are of course subject to further investigation with reference to the CSQT.

REFERENCES

Catell, R.B. The Scree Test for the Number of Factors. **Multivariate Behavioral Research**, 1966.

French, J.W., Ekstrom, R.B., and Price, L.A. **Manual for Kit of Reference Tests for Cognitive Factors**. Princeton, New Jersey: Educational Testing Service, 1963.

Guilford, J.P. and Hoepfner, R. **The Analysis of Intelligence**. New York: McGraw-Hill, 1971.

Iledan, B.R. Abilities Measured by the Test. **The College Scholarship Qualifying Test**. FAPE Review, Makati, Rizal: Office of Reports, FAPE, 1976.

Kaiser, H.F. **The Varimax Criterion for Analytic Rotation in Factor Analysis**. Psychometrika, 1958.

Parocha, A.A. Test Development. **The College Scholarship Qualifying Test**. FAPE Reivew, Makati, Rizal: Office of Reports, FAPE, 1976.

Pawlik, K. Concepts in Human Cognition and Aptitudes. In R.B. Cattell (Ed.) **Handbook of Multivariate Experimental Psychology**. Chicago: Rand McNally, 1966.

Royce, J.R. Epistemic Styles, Individuality, and World-View. **Research Bulletin**. Princeton, New Jersey: Educational Testing Service, 1973.

Smith, I.M. **Spatial Ability**. San Diego: Robert R. Knapp, 1964.

Vernon, P.E. **The Structure of Human Abilities**. New York: Wiley, 1950.

Witkins, H.A. Cognitive Styles in Learning and Teaching. In S. Messick and Associates (Eds.) **Individuality in Learning**. San Francisco: Jossey-Bass, 1976.

The Development of Occupational Norms for The Philippine Aptitude Classification Test

ESPERANZA C. BUEN
THOMAS E. DOHM

INTRODUCTION

The problem of proper utilization of manpower resources in the Philippines has been a growing concern in our country. Many of the graduates have difficulty in finding a job in their field of training whereas in other occupa-areas, there is a scarcity of graduates to choose from.

One of the possible reasons for the imbalance could be that students as a whole do not make realistic career choices. This is a finding based upon research on 300,000 students seeking admission to college (Dohm, 1974). Thus, one of the major reasons for the development of the Philippine Aptitude Classification Test (PACT) was to help rationalize career decisions among college-bound students. In addition to this, the test may also serve as a device to upgrade the placement process into jobs and to influence manpower planning in various types of companies and organizations.

The basic rationale behind the PACT Development Project is that aptitudes are required in different combinations and in varying degrees for successful performance in different occupations. These aptitude requirements can be determined by developing norms on different occupations. A college-bound student's performance in the test can be compared with these norms and judgment can be made as to which course or occupation his aptitudes are most suited. Likewise, a job applicant can be assessed in relation to the norms established for the jobs for which he is being considered.

The development of norms for different areas of training and employment takes a considerable amount of time and resources. The PACT Development Project (PACTDP) has been undertaking this task gradually. In 1976-1977, it started to develop norms for courses of training. Norms on 23 courses were established, 14 on academic/professional areas and 9 on vocational/technical areas. These were based on data gathered from graduating students of different colleges and vocational schools (PACT, 1977). These are now being used for the aptitude classification of high school students who have taken the PACT under the auspices of the Center for Educational Measurement (CEM).

In September, 1977 the PACTDP started to survey the business and industrial field for possible sample sources for the development of occupational norms. It contacted a central organization, the Philippine Chamber of Industries (PCI) which has many affiliate member companies or institutions. Interested companies that responded were informed of the general research plan and their involvement in it, as well as the benefits they could derive from their participation. Furthermore, preliminary inquiry was made in order to obtain specific information related to project requirements. When agreement was reached between the CEM and a company, the development of norms for selected occupations was formally started.

The purpose of this paper is to present the basic process by which the initial norms for jobs and occupations was developed for the PACT. Aside from discussing summary results for ten broad occupational groups, a more detailed analysis for one norms sample will be explained.

A major portion of the methodology in this study was patterned essentially after that used for the General Aptitude Test Battery (GATB). All instruments, however were produced locally. The key variables involved were occupational groups as the independent or classification variable and performance on the PACT as the dependent variable. Optimum combinations of factors on this test made up the norms for the occupational groups. Job performance or satisfactoriness of a worker as judged by his supervisor was the criterion against which the norms for an occupational group were validated.

Aside from setting limits to age, tenure, educational and physical qualifications, job satisfaction was used as an additional criterion for sample selection. This meant the worker's general attitude of contentment with his job and relevant features of the job setting. This was based on the theory that an individual's work adjustment at any point in time is defined by the concurrent levels of both his satisfactoriness and satisfaction (Dawis, 1964).

METHOD

Design

A disproportional stratified random sampling design was used, where the stratification variable was occupational group.

Subjects

The initial sample of 577 employees came from two industrial organizations. The ranged from 20-60 years of age, with at least one year of employment and had no physical handicap.

Instruments

The following forms were used in the study:

Philippine Aptitude Classification Test (PACT). This is a test designed to measure aptitudes for different occupations. It consists of three parts, two power test and one speed test. Seven aptitude factors were measured: Verbal Pilipino, Verbal English, Numeric, Spatial, Perceptual Speed, Induction, and Clerical Perception.

Personnel Data Form. This is a questionnaire which contains items pertaining to some demographic, educational, socio-economic and job status variables.

Job Attitude Survey. This form contains 80 Likert-type items regarding an employee's attitudes towards different aspects of his job such as job properties, interaction context and organizational policy. The form includes both English items and their Tagalog translations. It has a split-half reliability coefficient of .8235 ($P < .001$) and its internal validity coefficient with a general satisfaction score is .4842 ($P < .001$).

Performance Evaluation Form. This was used by supervisors to record the job satisfactoriness of employees under their supervision who participated in the study. This form was patterned after that used for the GATB (1970) norm development study. It has a reliability coefficient of .8200 ($P < .001$) and internal validity coefficient (with an over-all performance score) of .8274 ($P < .001$).

Procedure

Job Selection. After a formal agreement was arrived at between the companies and the testing agency (CEM), a brief description of the nature and objectives of each company was secured. A list of jobs was submitted and a number of them were selected according to the priorities of the CEM and the company and the adequacy of incumbents per job. More detailed specifications and selection requirements were obtained for these jobs. Broader occupational groupings were made for jobs which had very close similarities in specifications.

Sample selection. Once the final jobs were chosen, the employees to represent each group were randomly selected from the payroll list. A final list of examinees was prepared. In order to assure confidentiality each person was assigned a code number which he used throughout the study. Invitation to participate was made through the company coordinator.

Administration of Personnel Data Form, Job Attitude Survey and PACT Battery. The subjects were tested in groups using the standard manual for administration (PACT, 1977). To avoid interruption in work flow, examinees for one session did not have to come from the same job position or department. The rooms used were examined for adequacy of testing conditions. For each group of 35-40 employees, one examiner and one proctor were assigned. The instruments were administered in the following order: Personnel Data Form, Job Attitude Survey and PACT. This took four hours at the most.

Collection of Performance Data. The list of actual participants per department was sent to the respective heads who in turn handed over the particular names and code numbers to the direct supervisors. All supervisors or heads concerned were briefed on the rationale and the use of the rating form. They filled out the forms using employees' code numbers for identification. The accomplished forms were retrieved a month after the testing.

Analysis

Data obtained in this study were subjected to the following types of analysis:

Descriptive Analysis. Frequency distributions and descriptive statistics were computed for relevant demographic variables to see if the sample satisfied criterion requirements and to describe it more adequately. Furthermore, Job Satisfaction and Performance Rating Scores were computed and employees who scored Highly Dissatisfied or rated Poor in performance were eliminated from the group. The final sample was broken down into jobs or occupational groups.

Qualitative Job Analysis. The information about each occupational group was studied in relation with PACT aptitude factors. Judgment was made as to which factors were important to efficient job performance. Irrelevant factors were also identified. Reasons for the judgment were stated.

Quantitative Analysis. Statistics were obtained on each norm group based on their PACT factor scores and Performance scores. These were: Means, standard deviations, ranges of PACT factor scores and Pearson correlation coefficients between them and Performance scores. Stepwise multiple regression analysis was also done with PACT factors as the independent variable and Performance as the dependent variable. The F-ratios of regression variance were tested for significance.

Determination of Trial Norms. Three to four factors were considered as trial norms if they satisfied the following conditions:

a. a high mean score relative to other means obtained for the norm group.
b. a low standard deviation relative to the general sample and the other SD's obtained for the norm group. (SD = 100 is considered low).
c. a correlation with the criterion that is significant at least the .05 level.
d. a rating of "important" on the basis of job analysis.
e. a significant F-ratio on the stepwise regression analysis.

Determination of Final Norms. Cut-off scores were determined to separate the upper and lower 27%-30% of the group with respect to performance. Cut-off values for each of the trial factors were tentatively computed around their mean. Using these cutting values, the combinations of trial factors and trial cut-off values were correlated with the criterion. The phi coefficient (ϕ) was used as the measure of this relationship. The cut-off values of combinations with the highest significant ϕ coefficients were considered as final norms for the occupational group. The effectiveness of these norms was evaluated in terms of the proportion of qualifying and non-qualifying test scores in relation to the high and low criterion groups. Effective norms should be able to screen out at least 50% of the relatively poor workers (GATB, 1967).

RESULTS AND DISCUSSION

The initial occupational norms for the Philippine Aptitude Classification Test (PACT) were based on 547 employees from two industrial companies in Metro Manila. The sample represented various jobs or positions but may be classified under ten (10) broad occupational groups. These were the following: electrical engineer, mechanical engineer, administrative supervisor/assistant, foreman, bookkeeper, accounting clerk, general clerk, electrician, technician and lineman.

Table 1 shows that the sampling distribution met the criterion requirements for the size that should constitute a norms group. The largest group consisted of Administrative supervisors and assistants (71) while the smallest group was composed of Lineman, with only 35 workers. As recommended in GATB (1970), norms are never established on samples of less than 30 workers.

Table 1 Distribution of Subjects by Occupational Group

Occupational Group	Code No.	F	%
Electrical Engineer	00310	58	10.53%
Mechanical Engineer	00324	52	9.44
Administrative supervisor/ assistant	11910	71	12.88
Foreman	11930	54	9.80
Bookeeper	20100	55	10.16
Accounting Clerk	20200	65	11.98
General Clerk	21000	64	11.98
Electrician	76100	52	9.44
Technician	76200	41	7.44
Lineman	76520	35	6.35
Total		547	100%

NB. Code Nos. for occupational groups were obtained from the Philippine Standard Occupational Classification.

The sample also fulfilled the requirements for Age and Tenure as evidenced from statistics indicated in Table 2 below. The age range is from 21 to 56 years, with a mean of 33 and a mode of 31. The minimum number of years (1) was also the mode of the group. Although there were older workers (56 years old) with long tenure (24 years on the job), the group as a whole was relatively young and had not stayed very long in their present positions (Mean = 3.7 years).

With regard to educational background, the employees have finished the minimum level required of the norm groups, that is, at least a secondary edu-

Table 2 Selected Descriptive Statistics for Age and Tenure of Subjects

Variable (in years)	Mean	SD	Range	Mode
Age	33	7.5	21-56	31
Tenure	3.7	3.4	1-24	1

Table 3 Educational Attainment of Norm Group

Occupational Group	Educational Level					Total
	High School	Some College	Voc'l/ Tech.	College Degree	MA/MS Degree	
Electrical Engineer	0	1	0	57	0	58
Mechanical Engineer	0	1	0	51	0	52
Administrative Supervisor/ Assistant	2	19	0	48	2	71
Foreman	26	13	12	3	0	54
Bookkeeper	3	19	1	31	1	55
Accounting Clerk	2	19	1	43	0	65
General Clerk	5	29	4	25	1	64
Electrician	5	16	24	7	0	52
Technician	1	20	14	6	0	41
Lineman	23	7	3	2	0	35
Total	67 12,2%	144 26.3%	59 10.8%	273 49.9%	4 .8%	547 100%

cation (See Table 3). This was also the mode level for Foreman (26) and Lineman (23). The electricians and technicians had either finished some college or a vocational/technical training while the administrative employees as well as the clerks and bookkeepers either finished college or had attended a few years in it. Two hundred seventy-three (273) out of 547 or 49.9% of the total sample had attained a college degree. This was the level expected of the engineers and practically all of them (51 out of 52) satisfied this requirement.

It must be pointed out that the sample described above was the group of employees that remained after rejecting those who rated low in Satisfaction and Satisfactoriness (Performance). Out of the initial 577 employees who were tested, 30 or 5.2% were reported as **very dissatisfied** or **poor in performance** as measured by the Job Attitude Survey and the Employee Evaluation Form, respectively. Thus, the final sample of 547 workers consisted of those who were relatively satisfied and judged satisfactory in their jobs from the management's point of view.

PACT Profile of Norm Groups

From the series of analysis that was performed, a norm profile was set up for each of the ten occupational groups. This profile consisted of aptitude factors which were found most relevant to the group. The seven factors which were summarily analyzed were the following: Verbal Pilipino (VP); Numeric (N); Verbal English (VE); Perceptual Speed (PS); Spatial (S); Induction (I) and Clerical Perception (CP).

Table 4 shows the factors found relevant for the ten occupational groups. The first set of factors was considered important to each job after a qualitative analysis of specifications and requirements. The second set was the result of a comparison of means and standard deviations as well as tests of significance of the correlation between PACT factor scores and

Table 4 Relevant PACT Aptitudes Per Occupational Group as Determined by Qualitative and Quantitative Analysis

Occupational Group	Relevant PACT Factors		
	(1) From Qualitative Job Analysis	(2) From Comparison of M, SD and Pearson R	(3) From Regression Analysis
Electrical Engineer	N, VE, S, I	VP, I	VE, I, VP,
Mechanical Engineer	N, VE, S, I	VP, VE, I	VE, CP, S, I
Adm. supervisor/ assistant	N, VE, I, CP	VP, VE, I,CP	I, VE, CP
Foreman	I, CP	N, VE, CP	N, CP, VE
Bookkeeper	N, PS, CP	VP, N, VE, CP	N, VP, CP, VE
Accounting Clerk	N, PS, CP	VP, VE, PS, CP	VP, N, PS, CP
General CLerk	VE, PS, CP	VP, VE, I, CP	VP, VE, CP
Electrician	PS, S	VP, VE, PS, CP	VP, VE, S
Technician	PS, S, CP	VP, VE, S	CP, S, VP
Lineman	VE, PS, S, CP	VE, PS, N	PS, CP, VP

performance scores. On the third set, the aptitudes that were found most predictive of performance were determined by multiple regression using the stepwise technique. The factors used were those found relevant in the analysis 1) and 2). In this method the first predictor considered in the solution is the one which has the highest correlation with the criterion, and which subsequently contributes most to the variance in performance.

It is understandable that the set of aptitudes selected for each norm group would vary with the type of analysis that was used. The next logical step was to try out different combinations of the aptitudes that emerged relevant for validity against the performance criterion. The phi coefficient (∅) was used to indicate this relationship. Trial combinations of factors, based on the relevant sets in Table 4, as well as the trial cut-off scores per factor were tested with this statistic. The final optimum combination with their corresponding cut-off values consisted of the set which obtained the maximum ϕ. Table 5 shows this combination and the corresponding ∅. This is because from the set of trial norms, a series of crosstabulation analysis was done and the combination of aptitudes, with respective cut-off values (based on their means and SD's) which yielded the highest ∅'s were considered the final norms for the occupational group. In the particular sample, only eight (8) out of the ten groups had significant ∅ values. It seemed that the norms for Electrical Engineer and Foreman did not correlate strongly enough with the criterion. In all the rest of the groups, the norms were able to discriminate between employees who were highly satisfactory in their job performance and those who were relatively less satisfactory.

Table 5 Validity of PACT Aptitude Norms for Each Occupational Group

Occupational Group	Optimum Combination of PACT Factors (with cut-off values)	∅
Electrical Engineer	VE (483); I (589); VP (542)	.03315
Mechanical Engineer	VE (518); CP (425); S (441); I (482)	.35479*
Administration supervisor/ assistant	I (446); VE (520); CP (430)	.42322**
Foreman	N (343); CP (234); VE (374)	.17920
Bookkeeper	N (327); VP (532); CP (332); VE (401	.39135*
Accounting Clerk	VP (450); PS (399); CP (433)	.30232*
General Clerk	VP (576); VE (461); CP (462)	.39853**
Electrician	VP (495); VE (399); S (395)	.43819*
Technician	CP (411); S (343); VP (405); PS (277)	.39728*
Lineman	PS (233); CP (284); VP (437); N (342)	.50252*

*significant at p .05
**significant at p .01

The last step in the norm development was to test the effectiveness of norms by direct examination of the two-way tables in the crosstabulation analysis. Norms would be considered effective if they were able to screen out at least 50% of the low performers in a job group. Averaging out the data from the different groups with significant ∅'s, 57.12% of the poor workers did not achieve the minimum cutting scores on the aptitudes. Moreover, 80.59% of the workers who made qualifying test scores were good workers.

An Illustrative Case: Bookkeeper (20100)

A more detailed example of the development of norms for a particular occupational group will be treated here. The case we have chosen is the norm group of Bookeeper (code no. 20100).

Job Summary (obtained from job description). The bookkeeper is responsible for the preparation and summarization of journal entries, in the posting of journal vouchers and check vouchers in the bank control ledger, and details of deposit ledger and voucher register.

Experimental sample. Fifty five (55) employed workers.

Criterion. Supervisory ratings based on performance for at least six months on the job.

Qualitative analysis. The job analysis in Table 6 indicated that Numeric (N), Clerical Perception (CP) and Perceptual Speed (PS) appear to be important in the performance of the duties of Bookkeeper. The other factors were not that important but were not also considered definitely irrelevant.

Table 6 Job Analysis for Bookkeeper

Judgment Category	PACT Aptitudes						
	VP	N	VE	PS	S	I	CP
Important		✓		✓			✓
Irrelevant							

Aptitude	Rationale
Numeric (N)	Necessary in the preparation of journal entries and their summarization.
Clerical Perception (CP)	Important in posting entries on journals, vouchers, checks, ledgers and registers.
Perceptual Speed (S)	Helpful in summarizing journal entries.

Comparison of Mean, SD and r's. Table 7 shows the statistical results obtained for the experimental sample of bookkeepers.

The aptitudes with relatively high scores are Verbal Pilipino (VP), Verbal English (VE), Induction (I), and Clerical Perception (CP). Those with low standard deviations (SD=100) are Verbal Pilipino (VP), Verbal English (VE), Perceptual Speed (PS).

The data show that Numeric (N) correlates with the criterion at the .001 level while Verbal English (VE) and Clerical Perception (CP) had r's significant at the .05 level.

Multiple Regression. Results of the stepwise regression indicate that Numeric (N), Verbal Pilipino (VP), Clerical Perception (CP) and Verbal English (VE), in the order mentioned, contributed most to the variance in the criterion scores. The multiple R of .42172 was significant at the .05 level. It had a corresponding F-ratio of 2.75806 which was greater than the significant value at $P < .05$ with degrees of freedom: 4, 50.

The values of the regression coefficients indicate that if performance were to be predicted on the basis of the relevant factors an equation of the following form would be applied:

Table 7 Means (M), Standard Deviations (SD), and Pearson Product-Moment Correlations with Criterion (r) for PACT Aptitudes

Occupational Group: Bookkeeper (20100)			
Aptitude	**M**	**SD**	**r**
VP – Verbal Pilipino	549.16	66.44	.1420
N – Numeric	456.22	104.18	.4636***
VE – Verbal English	525.52	73.78	.2455*
PS – Perceptual Speed	457.06	96.08	.0935
S – Spatial	459.49	108.71	.1230
I – Induction	506.91	128.08	.0326
CP – Clerical Perception	497.39	114.55	.2466*

*** significant at $P < .001$
* significant at $P < .05$

N = 55

Table 8 Selected Statistics from the Stepwise Regression of PACT Aptitudes with Performance Scores for Bookkeeper

PACT Factor	Multiple R	B
N – Numeric	.40961	.02194
VP – Verbal Pilipino	.41773	.00554
CP – Clerical Perception	.42130	.00320
VE – Verbal English	.42172*	.00207
(Intercept)		16.85768

*significant at P .05; d.f.: (4,50)

N = 55

Y = .02194 x N score + .00554 x VE score + .00320 x CP score + .00207 x VE score + 16.85768

Summary of qualitative and quantitative analysis. Table 9 summarizes the results of the job analysis, comparisons of M, SD, and r's and the regression analysis.

Table 9 Summary of Qualitative and Quantitative Data for Bookkeepers

Evidence	Aptitude VP	N	VE	PS	S	I	CP
Job Analysis (Important)		✓		✓			✓
Relatively High Mean	✓		✓				✓
Relatively Low SD	✓		✓	✓			
Significant Correlation with Criterion		✓	✓				✓
Significant Regression variance	✓	✓	✓				✓
To be Considered for Trial Norms	✓	✓	✓				✓

Determination of Norms. Based on the qualitative and quantitative evidence, Aptitudes VP, N, VE and CP were selected for further consideration for inclusion in the test norms. Trial norms consisting of various combinations of three or four of these aptitudes were evaluated against the criterion by means of the phi coefficient. For this analysis, the criterion was dichotomized by placing 15 of the 55 workers or 27% in the Low Criterion Group and 27% also in the High Criterion Group. Results of the analysis showed that the best selective efficiency was obtained for the test norms with the following cut-off scores: N – 327, VP – 532, CP – 332, VE – 401.

Effectiveness of norms. Table 10 expresses a "Hit – Miss" table which shows the relationship obtained between norms with cutting scores given above and the dichotomized criterion. If an employee did not meet the cut-off score for at least one of the relevant factors, he was classified as having a "non-qualifying test score." The data indicate that 9 of the 15 relatively poor workers, or 60% of them, did not achieve the minimum scores established as cut-off for the recommended test norms. This means that this proportion would not have been hired if these norms had been used in the selection process. Moreover, 78.6% of the highly satisfactory workers would have qualified.

CONCLUSION AND RECOMMENDATION

The foregoing discussion has covered both a summary treatment of the development of PACT norms for ten occupational groups and an illustrative example for one specific occupation. The analysis involved seems to be rather lengthy but in view of the use to which the norms would be applied, such a method is deemed worthwhile going through.

For the purpose of using the norms for selection and placement, as well as for career decisions, two methods are recommended for the interpretation of examinees' scores. One may use a regression equation based on the coefficients that resulted from the regression analysis and the ∅-test for a particular occupation. This is the present classification

Table 10 Relationship Between Test Norms and Dichotomized Criterion for Bookkeeper

	Non-Qualifying Test Scores	Qualifying Test Scores	Total
High Criterion Group	11 (78.6%)	3 (21.4%)	16
Low Criterion Group	6 (40.0%)	9 (60.0%)	15
Total	17	12	29

ϕ = .39135

P/2 < .025

method used for junior high school examinees with respect to the PACT course norms. A second way is to compare the factor scores with the aptitude cutting scores for the occupation/position being considered. This will involve the use of the standard error of measurement (SEM). These two methods can be studied for comparative efficiency as well as economic feasibility. For the latter reason, it is recommended that not all the possible courses or occupations be open to comparison but certain other criteria be considered (interest or choice, course or job availability) during the process of aptitude classification.

With regard to the applicability of the newly-developed occupational norms certain reservations will have to be made. First of all, the lack of discriminating efficiency in the case of Electrical engineer (00310) and Foreman (11930) suggests that another validation study be made on perhaps a more extended sample or a different criterion measure. Thus the PACT as of now should not be sued in predicting those two occupations.

Efficiency is predicted for the other eight occupational norms if used in the direct selection or placement of workers in the companies that have provided the standardization sample, but when other companies wish to use the PACT norms for their own screening purposes, it is suggested that check studies or cross-validation be done on the established norms on at least a representative sample from their company. This reservation is made because any selected sample used for norms can only approach perfect representation of the entire population of workers in that occupation. Other factors such as the reliability and validity of the criterion tend to limit the degree of certainty that can be placed on their results. Therefore it would not be wise to accept the results of any one study as the "true" and "final" results, and it is advisable to conduct check studies to verify original findings.

Finally, since the PACT norms study is still at its initial stage and the validity used is only concurrent, more norms can be developed for other relevant occupations and predictive or longitudinal validity studies can be undertaken for these norms.

REFERENCES

Dawis, R. Theory of Work Adjustment. **In Minnesota Studies in Vocational Rehabilitation.** 1964, 15.

Dohm, T. Foreword, **FAPE Review**, Vol. 5 (4) January 1975.

General Aptitude Test Battery, Vol. 2. Washington, D. C.: US Gov't Printing Office, 1967.

General Aptitude Test Battery Manual, Sections III and IV. Washington D. C.: US Gov't Printing Office, 1970.

Philippine Aptitude Classification Test, Administration Manual. Makati: FAPE Publication, 1976.

Philippine Aptitude Classification Test. Course Profile Handbook. Makati: FAPE Publication, 1977.

Philippine Standard Occupational Classification. Manila: NEDA Publications, 1977.

Implications of the CEM Testing Program to Mathematics Education

THOMAS E. DOHM
ELIZABETH S. SAMIA

Two tests have been developed by the Center for Educational Measurement for third year secondary students in the Philippines. These are the Philippine Aptitude Classification Test (PACT) and the Diagnostic Tests.

The PACT is a classification instrument which measures the relevant aptitudes necessary for proper placement into occupational areas. By comparing a high school student's aptitude profile with that of 23 college and vocational course norms, the PACT "matches" each student's scores with each of the aptitude patterns of the course norms to help the student make a more realistic career decision.

The seven aptitude factors of the PACT are: (1) Verbal Pilipino (VP), the ability to understand the meaning of Pilipino words and to recognize relationships between them; (2) Numerical (N), the ability to perform arithmetic operations accurately; (3) Verbal English (VE), the ability to understand the meaning of English words and to recognize relationships between them; (4) Perceptual Speed (PS), the ability to recognize similarities and differences among series of letters, numbers, or figures quickly and accurately; (5) Spatial (S), the ability to recognize figures from a complex pattern and to construct them mentally; (6) Induction (I), the ability to discover a general rule or principle behind a series of numbers or figures; and (7) Clerical Perception (CP), the ability to observe differences in printed words, numbers, and punctuation, and to recognize the correct spelling of English and Pilipino words.

On the other hand, the Diagnostic Tests are designed to measure the common core of learning to which students are exposed to in their third year of high school. The battery includes tests in English, Geometry, Chemistry and Social Studies and follows the requirements of the revised secondary curriculum. For purposes of this paper, only the Geometry Diagnostic Test will be discussed. The Geometry Diagnostic Test is a 120-item, one-hour-and-thirty-minute, multiple-choice test that measures content areas and cognitive skills as defined by Bloom's Taxonomy of Cognitive Objectives and modified for a Philippine audience. The cognitive skills measured include knowledge, comprehension, application, computational and analytic abilities in the area of Geometry. Likewise, the content areas cover geometric phehemena, measurement, formal proof, and coordinate geometry. The test puts greater weight to the solution of practical quantitative problems, tests of computational skills and reasoning in a quantitative context rather than memorization of formulas and knowledge of techniques.

This paper will discuss the results on the PACT and the Geometry Diagnostic of Philippine students. Specifically, this paper will:

1. Present a profile of mathematics students who are in their last semester of college. The profile will include a discussion of curricular requirements, PACT test scores and their relationship to success in the mathematics programs, and other relevant student data;
2. Discuss the strengths and weaknesses of third year high school students on Geometry Diagnostic Test; and
3. Present the findings on the performance of third year high school students on select items of the PACT and Geometry Diagnostic Test in comparison with a criterion of expected performance based on responses from a sample of 39 high school teachers.

The first objective will help secondary instructors and counselors to help their students plan for a tertiary course in mathematics. The last two objectives will provide some feedback to secondary educators as to the relative strengths and weaknesses of Philippine third year high school students in mathematics.

The PACT College Norm Profile of Mathematics Majors

Sample Description

Sixty mathematics majors who were completing the final semester of their undergraduate curriculum in 1975 were included in the norm group. These students originated from six well-known public and private universities in the Greater Manila area. The schools are considered to have good mathematics programs and relatively large provincial enrollments.

Curricular Requirements

The total number of units required for graduation with a major in mathematics at the six universities was subdivided into eight curricular areas. For each of the curricular areas the percentage of total units required was computed to determine how much emphasis was required in each of the areas. This information should be useful to students prior to entering a mathematics curriculum.

The results of the curricular analysis are as follows:

Major Subjects (Math),	– 44.63%
Natural Sciences	– 14.22%
Social Sciences	– 14.59%
Philosophy	– 3.95%
Arts/Humanities	– 1.31%
English	– 9.65%
Pilipino	– 1.76%
and others	– 7.60%

PACT Test Results

The PACT college norms were based on a group of students from 23 college and vocational courses from approximately 25 tertiary institutions. Each of the PACT factor scores was converted to a scale with a mean of 500 and a standard deviation of 100. The performance of the mathematics normative sample is better on each of the seven factors as compared to the average college sample. The mean scores of the mathematics majors ranged from 550 for Verbal Pilipino to 630 for Numerical. (Refer to Table 1)

Further correlational analyses were done for the mathematics majors to determine the extent to which each of the PACT factor scores correlated with the grades in mathematics subjects as well as with the grades in all other subjects. Table 2 shows the results of the correlational analysis:

Table 1. Mean Scores Obtained by Mathematics Majors in the PACT Aptitudes Factors

PACT APTITUDE FACTORS		MEAN SCORES
VP	Verbal Pilipino	550
PS	Perceptual Speed	556
CP	Clerical Perception	584
I	Induction	592
S	Spatial	594
VE	Verbal English	607
N	Numerical	630

Table 2. Correlational Matrix

PACT APTITUDE FACTOR	GPA-MATCH SUBJECTS	GPA-ALL SUBJECTS
Verbal Pilipino	–0.25*	–0.26*
Numerical	0.22*	0.21
Perceptual Speed	0.45**	0.45**
Verbal English	--0.06	0.03
Spatial	0.24*	0.28*
Induction	0.25*	0.27*
Clerical Perception	0.15	0.15

*Significant at .05 level
**Significant at .01 level

Interestingly enough, success in mathematics subjects as well as in college as a whole correlated negatively with verbal ability in Pilipino. This finding should have important implications on the Philippine bilingual program. Success in mathematics subjects seems to be most related to aptitudes for perceptual speed, induction, spatial, and numerical ability. Likewise, success in college as a whole for the mathematics majors depends on perceptual speed, spatial, and induction aptitudes.

Other interesting PACT Student Data Include the Following:

1. When asked to choose between English and Pilipino, 80% answered they could **read** better in English, 65% could **speak** better in Pilipino, and 71% could **write** better in English. Furthermore, 83.3% said they would prefer to take tests in English rather than in Pilipino.

2. From a list of given reasons for choosing their course, students responded as follows: 43.3% are interested in the course they chose, 28.3% think that their course choice promises better pay, 21.7% think that they have the ability to finish the course, 3.3% answered that others influenced their decision, to take the course 1.7% considered their course choice to be short and 1.7% had other reasons.

3. When preparing for lessons in major subjects, 56.7% answered they have an easy time, 40.0% find difficulty and 3.3% have a very difficult time.

4. When asked whether they would be happier in another course, 21.7% strongly disagreed, 56.7% disagreed and 21.7% agreed.

5. Finally, when asked to choose their course given another chance 35.0% would choose the same course, 51.7% would choose a different course but in a similar field and 13.3% would choose a completely different course.

Other interesting PACT Student Data Include the Following:

	Language Preference English	Pilipino
READ BETTER IN	80.0%	20.0%
SPEAK BETTER IN	35.0%	65.0%
WRITE BETTER IN	71.7%	16.7 %
LIKE TO TAKE TEST IN	83.3 %	16.7%

Reasons for Choice of Course:

I'M INTERESTED	43.3%
IT PROMISES BETTER PAY	28.3
I HAVE ABILITY	21.7
OTHERS WANT ME TO	3.3
SHORT TERM	1.7
OTHERS	1.7

Preparing Lessons in Major Courses:

EASY	56.7%
DIFFICULT	40.0
VERY DIFFICULT	3.3

When asked if they would be happier in another course:

STRONGLY DISAGREED	23.7%
DISAGREED	56.7
AGREED	21.7

Finally they answered that, if they had a choice, they would take:

SAME COURSE	35.0%
DIFFERENT COURSE BUT SIMILAR FIELD	51.7
COMPLETELY DIFFERENT	13.3

The implications of these findings are startling and should not be taken lightly. Apparently, stronger career guidance programs are needed at the secondary level. In summary, those who are working at the secondary level can help their students **plan** for a tertiary program in mathematics by taking note of the following:

1. Mathematics majors, in general, have higher aptitudes in the seven PACT factors than the average group of students in college.
2. The assumption that all that is needed for success in the mathematics curriculum is numerical facility is a myth. Perceptual speed, induction, and spatial aptitudes are the best predictors of success in mathematics programs. Verbal Pilipino negatively correlates with college success in mathematics possibly because the teaching of most mathematics subjects is done in English.
3. Surprisingly, a relatively large percentage of mathematics students are not satisfied with their courses. This finding needs to be explored further and may partly stem from the lack of proper career guidance at the secondary level.

The Geometry Diagnostic Test

The Geometry Diagnostic Test is most useful in determining the strengths and weaknesses of students at the **individual, institutional,** or at the **regional** level. However at the national level, the reporting of strengths and weaknesses is problematic due to the norm-referenced nature of the test. Since total scores and subscores are standardized to have a national mean of 500 and a standard deviation of 100, no subscore differences will occur at the national level due to the standardization and norming procedures. Although initially, the test was designed to be a criterion-referenced test, this plan was changed when the item judges could not effectively translate criterion measures at the item level. This was not due to any lack of ability on the part of the judges but rather to the perplexing problem of not being quite able to determine whether or not an item was difficult because the criterion it was supposed to measure was difficult or whether the item was difficult because of the particular language and terminology used in the item. Thus, the same criterion could be measured by two items of varying difficulty. Likewise, a comparison of subtest raw scores could not be directly compared for the national sample due to inherent differences in difficulty levels of the items in the various subtests. Because of these technical difficulties, indirect methods of determining strengths and weaknesses were made, namely:

1. Comparison of subtests' item rejection rates to determine the subtests' relative strengths and weaknesses; and
2. Comparison of actual performance of examinees with expected performance based on judges' ratings at the **item** level rather than at the subtest level. The reader could likewise make an assessment as to whether or not his own criterion for student performance had been achieved.

If one employs the first method, one makes several assumptions. If we have a group of items which are considered acceptable in terms of correctness of grammar, (add other criteria) etc., we may pretest or try out these items on a group of students. Items which are too difficult will be rejected if norm-referenced rules are applied to item selection. However, items which are rejected because of difficulty may be assumed to be due to the difficulty experienced by students in achieving the criteria (since grammar and other factors are acceptable). Therefore, the greater the percentage of items rejected in a particular subtest, the more difficult the criterion is assumed to be, and correspondingly this is where the students' weakness lie.

The following table shows the results of 8,660 students from Greater Manila, Central and Southern Luzon who took the February, 1976 pretest of the Geometry Diagnostic Test.

Table 3. Frequency of Rejected Items from the 1976 Pretest of the Geometry Diagnostic Test

	Number of Items Pretested	Number of Items Rejected	Percent of Items Rejected
I. CONTENT AREAS			
A. Geometric Phenomena	205	99	48.29
B. Measurement	123	74	60.16
C. Formal Proof	91	25	27.47
D. Coordinate Geometry	91	20	24.69
II. COGNITIVE SKILL AREAS			
A. Computation	156	72	46.15
B. Comprehension	144	59	40.97
C. Interpretation and Application	99	39	39.39
D. Analysis	98	47	47.96

Based on this indirect method of analysis, the weakest to the strongest content areas of the students are Measurement, Geometric Phenomena, Formal Proof, and Coordinate Geometry. Likewise, the weakest to strongest cognitive skills are Analysis, Computation, Comprehension. Interpretation and Application.

Furthermore, using the above method for the four CEM Diagnostic Tests, one finds that it is only in the Chemistry Diagnostic Test wherein the students are weaker than in Geometry Diagnostic Test. Students performed better in the English and Social Studies Diagnostic Tests than in the Geometry Diagnostic Test.

For the second method, select items were drawn from the 1975 PACT pretest, the 1976 Geometry Diagnostic Test, and the 1977 Geometry Diagnostic Test. Subsequent forms of these tests which are currently being administered nationwide no longer use the select items included here (See Appendix A).

Expected criterion of performance varies from educator to educator. Therefore, an attempt was made to determine the mean expected item criterion from a group of judges. This means estimate of expected item criterion performance might be used to evaluate the actual performance of students, and subsequently of the system as a whole. Thirty-nine judges were selected to estimate for each item the proportion of students at the end of their third year of high school who would answer the item correctly. The judges were told that the students came from a nationwide sample of public and private high schools from both the urban as well as the rural areas. Twenty-one of the judges taught high school in the Greater Manila Area whereas the remaining 18 taught in the provinces. All judges taught at the secondary level except one who was a tertiary instructor. The mean number of years of teaching experience was 6.1. Thirty-six of the judges were female, three were males. Lastly, 21 came from public schools, 13 from private sectarian schools, and 5 from private non-sectarian schools.

Each of the 24 items were administered to the Judges (Group D) and at least one of the following groups:

Group A – The nationwide November, 1975 pretest sample of the PACT. This includes 16,108 students from 76 schools.

Group B – February, 1976 pretest of the Geometry Diagnostic Test. Eight Thousand Six Hundred Sixty students from 104 schools in the Greater Manila Area, Central and Southern Luzon comprise the sample.

Group C – Nationwide February, 1977 Geometry Diagnostic Test examinees of the FAPE Centralized Testing Program. This includes 8,099 students. It should be noted that since the FAPE Centralized Testing Program is a voluntary program, there is a tendency for only the better students from the better schools to take the tests.

All of the student groups mentioned above came from both public and private schools.

For each item, the following table presents the actual proportion of students who answered the item correctly and the corresponding mean estimate by the judges of the expected proportion of students who would correclty answer the item.

Table 4. Difference between the actual proportion of student group answering an item correctly and the judges mean estimated proportion of students who would correctly answer

Item	Group	Actual Proportion of Student Group Answering Item Correctly	Mean Estimated Proportion From Judges (Group D)	Difference
1	(A)	0.25	0.58	0.33
2	(A)	0.23	0.68	0.45
3	(A)	0.27	0.82	0.55
4	(A)	0.16	0.72	0.56
5	(A)	0.19	0.55	0.36
6	(A)	0.21	0.69	0.48
7	(A)	0.22	0.71	0.49
8	(A)	0.19	0.82	0.63
9	B	0.37	0.66	0.29
	C	0.51	0.66	0.15
10	B	0.37	0.71	0.34
	C	0.49	0.71	0.22
11	B	0.49	0.79	0.30
	C	0.57	0.79	0.20
12	B	0.26	0.57	0.31
	C	0.36	0.57	0.21
13	B	0.36	0.48	0.12
	C	0.44	0.48	0.04
14	B	0.62	0.64	0.02
	C	0.56	0.64	0.08
15	B	0.35	0.64	0.29
	C	0.40	0.64	0.24
16	B	0.34	0.49	0.15
	C	0.34	0.49	0.15
17	B	0.35	0.50	0.15
	C	0.46	0.50	0.04
18	B	0.37	0.53	0.16
	C	0.51	0.53	0.02
19	B	0.31	0.61	0.30
	C	0.31	0.61	0.30
20	B	0.34	0.60	0.26
	C	0.43	0.60	0.17
21	B	0.41	0.73	0.32
	C	0.41	0.73	0.32
22	B	0.35	0.65	0.30
	C	0.36	0.65	0.29
23	B	0.44	0.60	0.16
	C	0.43	0.60	0.17
24	B	0.31	0.50	0.19
	C	0.31	0.50	0.19

From the above table, it is interesting to note that **in no case were the estimates** of the judges **lower** than the actual performance of third year high school students. Thus the judges had overestimated performance on all items. In fact, the mean difference in proportion between the judges and Group A (PACT) items was 0.48. The mean differences between Group B and the judges was 0.23 and the mean difference between Group C and the judges was 0.17.

If the judges' criteria were set to evaluate mathematics education as a whole, then the system has failed in meeting the expected criterion performance. The author will leave it up to the reader to estimate his own expected item performance and compare his own expectations with that of the actual performance of Philippine students. Have your own criterion been met? If not, the author will leave the implications of this to the reader. For after all, it is the Philippine mathematics educator who must search for causes and ultimately the solutions to the improvement of mathematics education in this country.

Appendix A

Questionnaire Administered to Judges (Group D)

NAME: ______________________________

Institutional Affiliation: ______________________________

Address: ______________________________

Please check: ☐ Public ☐ Private-Sectarian

☐ Private Non-Sectarian

Position: ______________________________

At what level do you teach?

☐ Elementary ☐ High School ☐ College

How long have you been teaching? ______________

Highest degree completed: ______________

Field of Specialization: ______________

DIRECTIONS:

The following items are intended for third year high school students in the Philippines, both for the public and private schools in the urban areas as well as the remote rural areas. The items will be administered at the **end** of the **third** year of high school. Please estimate that proportion of these students will correctly answer the item. Your answer should be accurate up to two decimal places. For example, if you expect 35% of the nationwide sample of third year students to answer it correctly, please place **0.35** next to the corresponding item. Note that an asterisk is placed by the correct answer. Thank you.

_____ 1. 625 – 64 – 100 = (A) 8 (B) 28 (C) 6 (D) 17 *(E) None

_____ 2. $\frac{1}{2} - \frac{1}{6} \times \frac{3}{5} =$ (A) $\frac{1}{5}$ (B) $\frac{2}{5}$ (C) $\frac{5}{9}$ (D) $1\frac{3}{5}$ *(E) None

_____ 3. $\frac{4}{9} \times \frac{5}{6} \times \frac{1}{3} =$ *(A) $\frac{10}{81}$ (B) $\frac{1}{2}$ (C) $\frac{3}{4}$ (D) $\frac{5}{9}$ (E) None

_____ 4. $\frac{2}{3} - \frac{5}{12} + \frac{1}{4} =$ *(A) $\frac{1}{2}$ (B) $\frac{3}{4}$ (C) $1\frac{1}{4}$ (D) $\frac{2}{3}$ (E) None

_____ 5. 36 + 169 + 324 = (A) 81 (B) 47 (C) 91 *(D) 37 (E) None

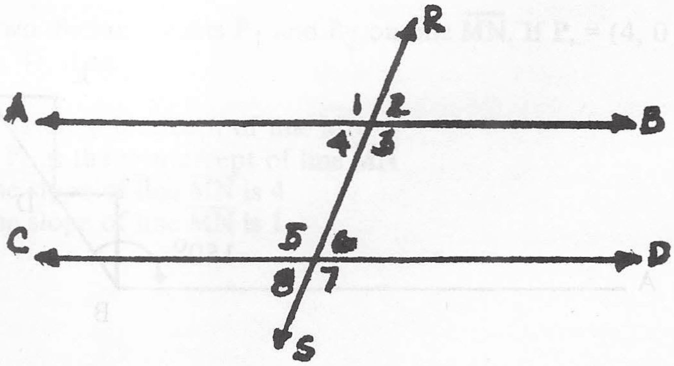
_____ 6. $\frac{1}{8} + \frac{3}{4} - \frac{5}{8} - \frac{1}{2} =$ (A) 1 (B) $\frac{3}{8}$ *(C) $\frac{1}{4}$ (D) $\frac{1}{2}$ (E) None

_____ 7. $\frac{5}{6} + \frac{3}{4} + 1\frac{1}{2} =$ (A) $2\frac{1}{3}$ (B) $1\frac{7}{10}$ (C) $3\frac{3}{4}$ *(D) $3\frac{1}{12}$ (E) None

_____ 8. $\frac{1}{3} \times \frac{1}{4} \times \frac{1}{5} \times \frac{2}{3} =$ (A) $\frac{5}{18}$ (B) $1\frac{4}{5}$ (C) $\frac{13}{17}$ *(D) $\frac{1}{90}$ (E) None

_____ 9. Given: A, B, and C are collinear points x, y and z, respectively;
If $x + y > x + z$, then (A) $y < z$ (B) $x \leq z$ *(C) $y > z$ (D) $x > z$

_____ 10. If $\overleftrightarrow{AB}$ is parallel to $\overleftrightarrow{CD}$ with $\overleftrightarrow{RS}$ as the transversal and M $\angle 3 = 134^\circ$, then m $\angle 6$ is



(A) 56° (B) 66° (C) 36° *(D) 46°

_____ 11. ABCD is quadrilateral. The sum of the angles A, B, C, D is

(A) 720° (B) 540° *(C) 360° (D) 180°

_____ 12. If one side of a square is three times as long as a second square, then the area of the first is how many times the area of the second?

(A) four (B) two (C) eight *(D) nine

_____ 13. If the number of degrees in ∠A of △ABC is represented by 5x + 25, in ∠B by 15x – 13, and in an exterior angle at C by 10 x + 54, the triangle is

*(A) Scalene (B) right (C) equiangular (D) obtuse

_____ 14. The perpendicular from the center of a circle to a chord divides the chord into

*(A) two congruent parts (B) two parts with ratio 3:4
(B) two parts with ratio 4:5 (D) two unequal parts with no definite ratio

_____ 15. Given the points A(3, -5), B(1, 1) and C(-4, 6), what can you say about the three points?

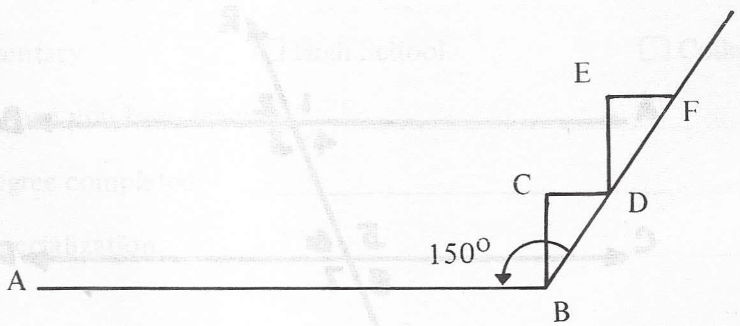
(A) C is between A and B (B) A is between B and C
*(C) The three points are non collinear (D) B is between A and C

_____ 16. What figure is the locus of the vertices of all isoceles triangles that can be constructed on line segment AB as a base?

(A) a log (B) the altitude to a log
(D) the base (*C) the altitude to the base

_____ 17. The timber BF supports a stairway and makes an angle of 150^{o} with the floor AB. How many degrees must CDB have so that CB AB and CD AB?

(A) 45^{o} *(B) 30^{o} (C) 60^{o} (D) 25^{o}



_______18. The midpoint of the hypotenuse is the circumcenter of a $30^o - 60^o - 90^o$ triangle. This is true firstly because

*(A) The median to the hypotenuse is equal to one half of the hypotenuse
(B) 1/2 of the hypotenuse is equal to the radius
(C) All the vertices will determine a circle
(D) The radius of the circle are congruent

_______19. In what kind of triangle do the altitudes meet at one of the vertices of the triangle?

(A) equilateral triangle (B) isosceles triangle
*(C) right triangle (D) scalene triangle

_______20. Two polygons having equal areas are congruent. The hypotenuse is

(A) If there are two polygons . . .
(B) If two polygons are equal . . .
(C) If two polygons are congruent . . .
*(D) If two polygons have equal areas . . .

_______21. If $\angle x = \angle y$ and $\angle y = \angle z$, then $\angle x = \angle z$, this statement is a direct application of which of the following properties

(A) Addition (B) Reflexibility (C) Symmetry *(D) Transitivity

_______22. If $m\angle A + m\angle B = 180^o$, and $m\angle A + m\angle C = 180^o$, then

*(A) $\angle B = \angle C$ (B) $\angle m \angle B + m\angle C = 180^o$ (C) $\angle A = \angle C$

(D) $m\angle B + m\angle C = 180^o$

_______23. The midpoint of the segment joining points (0, 0) and (−8, 0) is point:

*(A) (−4, 5) (B) (−4, −5) (C) (4, 5) (D) (4, −5)

_______24. Given two distinct points P_1 and P_2 on line $\overline{MN}$. If P, = (4, 0) and P_2 = (0, 4), then

(A) P_1 is the y-intercept of line MN
*(B) P_2 is the y-intercept of line MN
(C) The slope of line $\overline{MN}$ is 4
(D) The slope of line $\overline{MN}$ is 1.

Selecting Tests for Schools

ALEXANDER A. CALATA

Many schools in the Philippines select and administer tests without much purpose, plan or preparation. Schools that claim they have a testing program actually are saying they have a collection of assorted tests – foreign and local, mostly foreign – that they administer to certain grade or year levels and whose validity to the school's instructional objectives may be difficult to ascertain. Further, the understanding of these tests by the school administrators and faculty alike may not be perfect, an understanding that impairs more than clarifies their appreciation of student abilities and its implications to teachings.

The selection of tests is one of the most neglected aspects in the planning of a school's testing program. To many in the school system, the task of choosing the tests can be handled by almost anybody who is in education. And within the educational system, a lot of emaciated suggestions on tests are apparently frequently given, and, judging from the results, are faithfully discharged by schools.

On the other hand, the careful selection of tests for a school's needs can greatly influence the success of its educational objectives. When those in charge of this important task seriously try to relate tests to students, teachers, curriculum and school system, an institution's testing program can begin to produce positive results.

What, then, should be done?

II

In selecting tests for a school's testing program, **a first consideration is whether the tests will do the job you want done by them, and well.** It is obviously the most sensible thing to ask in the first place, and yet many fail on this count. Standardized (those prepared by specialists) and teacher-made tests will both serve the purpose of the school's general educational objectives and the specific objectives of classroom instruction. These objectives are important in defining the evaluation approach of a school. Although schools will vary in the way they state their objectives, they will however agree on some common core of abilities. For example, schools will always converge on the need to assess skills in reading comprehension and arithmetic computation. What is apparent is the ease with which school administrators and teachers identify areas to be measured. The problem seems to be what tests to use.

III

Now, tests are many things to many people. The whole educational system may view them as bases for systems-wide policies (e.g. the NCEE, SOUTELE) School systems may look at them as instruments to help determine academic strengths and weaknesses of students for possible curricular realignment and remediation programs (e.g. the diagnostic tests of the Notre Dame Educational Association schools in Southern Mindanao). Within the school system itself, tests will serve varying needs. A teacher regards them as an instructional aid. A guidance counselor interprets them as some concrete guideposts for defining a student's career path. An admission officers relies on tests for decisions on acceptance and grouping of incoming students. Within this welter of needs, what is important is for a school to define the areas that should be assessed and then find out what tests are needed to do this evaluation.

IV

Once the areas in which testing is to be done have been adequately defined, the school can **review the available tests in the market.** Information on tests can be had from catalogs which test publishers issue from time to time. Catalogs are listings of tests sold by companies. They contain information on the title of each test, the test's author(s), the appropriate target group for which the test is addressed, its manner of purchase and use, test administration mechanics, design of test materials, prices and other similar information. **Caveat emptor.** Most of the catalogs in the local market are from American test publishers. Catalogs on Philippine educational tests are difficult to come by, which is also one way of saying that there are only a few local tests available in the market. Outside of government, no organized national educational test development program is evident in the country except that of the Center for Educational Measurement (CEM). However the CEM does not sell its tests in the tradition of test publishers in the United States like the Psychological Corporation, McGraw Hill Book Company, or Science Research Associates. The best known commercial distributor of foreign educational tests in the local market is the Philippine Psychological Corporation. A more comprehensive and regular listing of major United States test publishers is made by the Educational Testing Service at Princeton, New Jersey, a non-profit educational testing agency like CEM. The ETS lists are given free of charge to interested schools. The most comprehensive source of test information and critique are the editions of Mental Measurements Yearbook by Oscar K. Buros. The Yearbook lists most of the current tests in print and contains information on each test's validity, reliability, and norms. It should also be pointed out that in some schools in the Philippines, particularly in the graduate schools, students are known to have developed educational tests for specific grade and year levels as part of their thesis work. In addition, many colleges and universities publish their research findings on the tests they use for different purposes. These findings as well as selected thesis data can help those in search of tests make conclusive decisions about their use or rejection.

V

Which brings us to the same problem of having the tests but not knowing what they can do. What they can do can be determined **when you examine the tests that you have or that you wish to obtain** according to their (1) quality and (2) ease of administration and scoring. The quality of a test essentially means its validity and reliability. In achievement tests, content validity must be extensively considered. This means that the content of a particular test should relate significantly with the content and emphasis of a particular course. Whether teacher-made or standardized, an achievement test should reflect what is taught in the classroom. For aptitude tests, of utmost importance is predictive validity, i.e., how well scores on a test correlate with some standard of future success. For example, in making a selection, a high school may wish to know the correlation between scores on the CEM-developed College Scholastic Aptitude Test (CSAT) and college freshman grades. If they correlate significantly, at least with a correlation coefficient equal to .30, then this could lead to a decision on the use of the test. Reliability is a consistency index. It says in effect that a reliable test yields the same results in measuring whatever it is supposed to measure. Reliability coefficients of tests are usually given in test manuals and therefore makes easy the task of a school in search of a reliable instrument. It should be noted that tests are never perfectly reliable. It is sufficient perhaps to say that the teacher or the school should seek a standardized test whose reliability is as high as possible. In general, most standardized tests published for school use have reliability coefficients of at least .80 in the population for which it is designed. What is important is for the school to interpret a test's reliability coefficient in terms of the groups of students on which it is based.

Other technical considerations that should be made concern the kinds of norms provided and the evidence of careful development and research. Of course, a careful study of all these technical attributes is not often feasible in the Philippine school setting, given its limitations, in which case the selection of the tests on the basis of these attributes may be based on recommendations of external experts.

Outside of the technical considerations, the tests should be selected on the basis of ease of administration, scoring and even the speedy and steady availability of other related services that test publishers can offer test users. Such services would include processing of results, research and analysis studies and interpretation of results.

VI

In the scrutiny of tests to be used, **note that there are some tests which every student should take and other tests which will be of use only for specific needs.** Needless to say, a decision on which test to use universally or selectively will be ultimately based on the user's purpose. A group test of mental ability, for instance, should be taken by all students, whereas a special aptitude (e.g. music) need not be administered to all Aptitude tests may be helpful if given to high school juniors or seniors when systematic focus on future vocational plans begins, but they are seldom of value to elementary school pupils since elementary schools usually follow a universal curriculum.

VII

In this connection, the school may want to **consider the wisdom of participating in a centralized testing program that is available to schools.** This program may be associational/regional (NDEA, Educational Testing Center) or national

(CEM). One big advantage of joining a testing program of this type is that the school is freed from the task of developing or selecting its own tests at the same time that such programs make available dependable norms for schools comparable to the local school. A comparison of these associational or national norms will help the school test consumer in analyzing the strong and weak areas of its own programs. It will also provide the school with data on comparative ability performances of students. When a school decides on a centralized testing program, however, it should first see to it that what the test contains agrees with its objectives and curriculum and provides adequate measurement of the entire range of ability of the students in the local school. An ideal and realistic arrangement perhaps would be a program of test complementation where standardized tests from major associational or national testing programs and school-developed instruments can be coordinated and integrated to cover, in an adequate fashion, the objectives, curriculum and ability range of the students.

VIII

This brings us to the question of who should select the tests. It is important to **have the tests represent all the critical claims and interests of administrators, counselors, the testing department (where there is one), and most important. the classroom teachers.** This makes sure that the tests are faithful to school objectives and needs and generally endorsed by all. The best way to do this is constitute a committee to be composed of these sectors. Where tests are departmentalized, i.e., they are developed by subject matter departments, coordinators and supervisors should involve their department teachers in all phases of test development work–from the preparation of test specifications and the writing and assembly of items to item analysis and test improvement.

While it is, as a rule, highly desirable to have a common testing program, there may be circumstances under which the choice of tests should be left to individual schools within the system. Perhaps the primary consideration in joining a common program is whether the range of student abilities in the different schools is similar and whether there is a common of core educational objectives which guide the various schools in the system.

IX

In summary, the selection of school tests will have to consider the following:

1. Define the school's educational and instructional objectives and find out whether the use of certain tests will achieve these objectives satisfactorily.
2. Based on a definition of educational and instructional objectives, determine the specific areas to be measured.
3. Learn what tests are available.
4. Choose the tests that will meet needs defined in nos. 1 and 2.
5. Examine tests according to their technical qualities and administrative requisites.
6. Select tests that reflect the thinking of administrators, teachers, counselors and testing departments.
7. Consider obtaining tests from associations or nationally planned testing programs. Where this is not sufficient to meet particular circumstances, select other appropriate tests.

The need for a careful assessment of school and student performance cannot be overemphasized. Amid the big number of tests available and their evident impact on the education of the young, schools should exercise the most thoughtful attention to selection and provision of the best testing materials.

REFERENCES

Katz, Martin. **Selecting an Achievement Test: Principles and Procedures.** Princeton, N. J.: Educational Testing Service, 1973.

Lennon, Roger T. **Selection and Provision of Testing Materials.** New York: Harcourt Brace Jovanovich, Inc., 1962.

Linden, D. and Linden, K. W. **Tests on Trial.** Boston: Houghton Mifflin Company, 1968.

Lyman, Howard B. **Intelligence, Aptitude and Achievement Testing.** Boston: Houghton Mifflin Company, 1968. "The Test Manual," in **Test Scores and What They Mean.**

A Pilot Study on the Implementation of a Model for Evaluating School Guidance Programs

MARIA CRISTINA CORPUS

With the promulgation of the revised secondary curriculum in 1973, there came a directive from the Education Department to "develop guidance programs where there are none and strengthen those which are already existing" (The Revised Secondary Education Program, May, 1973). Clearly, this underscored the significant role guidance was to assume in the light of new thrusts in secondary education. Whereas in the past, guidance was left to the discretion of individual institutions, by 1973, it became mandatory in all secondary schools across the nation. Perhaps the single, most compelling reason for the importance ascribed to guidance was the finding that its lack or absence in the early high school years often contributed to poor occupational choices among large numbers of graduating students.

With guidance emerging as a vital component in our educational system, counselors of necessity must engage in the systematic planning and implementation of guidance programs. Ultimately, the total impact of these programs needs to be assessed.

When viewed within the local context where guidance has now become one of the delivery systems for accomplishing national education objectives, evaluation is by no means a desideratum. Through evaluation, information on whether guidance programs are helping students attain national education objectives are made available on the local level, to guidance personnel concerned with the development of school guidance programs; and on the national level, to educators who are engaged in educational planning.

But to conduct evaluation, the evaluator must have access to a wide variety of methods and tools. Without them. evaluation is not possible. Today there are several evaluation approaches at the disposal of the counselor. One is the CIPP (Context, Input, Process and Product) Model. Devised by Stufflebeame (1968) and used for the evaluation of educational projects in the United States, it became necessary to develop a local adaptation of the CIPP to enchance its applicability in the Philippine context. This research study was thus initiated to determine the workability of the locally adapted CIPP model in selected schools on the secondary level.

Following its prototype, the pilot model identifies four phases in guidance program development, namely, planning, structuring, implementing and recycling. Each, in turn, is served by four kinds of evaluation: context, input, process and product. Thus, context evaluation serves program planning, input evaluation serves program structuring, process evaluation serves program implementation and product evaluation serves program recycling (cf. Figure 1).

Each of the four phases in developing guidance programs, as indicated above, is accompanied by a corresponding type of evaluation which provides for continuous and necessary feedback.

In accomplishing the four strategies for evaluation, the pilot Model prescribes a series of activities. Accompanying these are materials and instruments to facilitate the completion of the required activities. (cf. Figure 2).

The preliminary tryout of the pilot Model involved three secondary schools with small, medium and large student populations. In these schools, the counselor-student ratio ranged from 1:400 to 1:1,600. Implementing the pilot Model in their respective schools were guidance teams composed of the guidance coordinator and a counselor. A total of 344 students from these three schools comprised the pilot student sample.

Results of the pilot study were substantial and encouraging. Of the activities prescribed for implementing the four evaluation strategies, only two were not completed in two of the pilot schools. These activities focused on the development of the community and the school profiles, a specification under context evaluation. The succeeding activities were implemented in all the pilot schools, however. These included: analyzing counselor time expenditures, identifying stu-

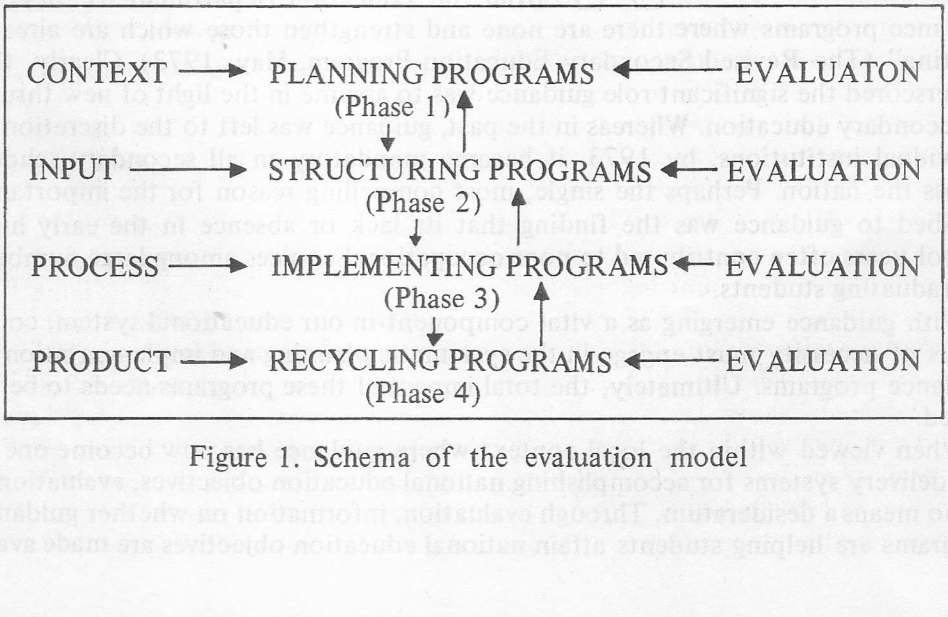


Figure 1. Schema of the evaluation model

	ACTIVITY	MATERIALS/INSTRUMENTS
CONTEXT	Analyze the real-life environment	Developing the community profile
		Developing the school profile
		Time analysis of counselor functions
		Identifying student needs
	Establish program goals	Developing guidance program goals
INPUT	Establish program objectives	Developing guidance program objectives
	Select appropriate guidance strategies	Determining workability, diffusibility, and possible benefits of selected strategies
	Pre-assess student's behavior prior to applicationof guidance strategies	Student Needs Assessment Survey (pretest)
PROCESS	Record program activities	Guidance program log
	Monitor program implementation	Notes on monitoring program implementation
	Post-assess students' behavior after application of guidance strategies	Student Needs Assessment Survey (postest)
PRODUCT	Evaluate and report program outcomes	Notes on product evaluation

Figure 2. Activities, materials, and instruments for evaluating guidance programs

dent needs, and establishing program goals (for context evaluation); developing behavioral objectives, and selecting appropriate guidance strategies (for input evaluation); maintaining the Guidance Program Log, and monitoring program implementation (for process evaluation); and evaluating/reporting program (for product evaluation).

Despite constraints of time, resources, and the lack of support from school administrators in one pilot school, the guidance teams reported several positive outcomes resulting from their use of the pilot Model. Some of these outcomes were: there was a noted increase in the number of students seeking counselor assistance after they were pretested on the Student Needs Assessment Survey; the model enabled counselors to fit guidance activities into an input-output system; through the Guidance Program Log, school administrators were furnished with a written record of what the counselor did, thus, increasing her accountability, the Student Needs Assessment Survey provided counselors with a criterion measure for evaluating the effectiveness of their programs; the model assisted counselors in the writing of behavioral objectives; and finally, counselors became more aware of the importance of learning about the real-life environment before proceeding with the development of guidance programs.

Further proof of the favorable impact generated by the study was obtained from two schools where a more thorough implementation of the pilot Model was planned the following school year.

The study likewise confirmed that "Developing the Community and School Profiles," materials which were prescribed in the context evaluation phase, have to be revised. Also, that the Student Needs Assessment Survey undergo further refinement. More field tests are therefore recommended not merely to improve the pilot Model but to establish, with greater confidence, its workability in other local guidance settings.

REFERENCES

Cook, D. **A Systems Approach to the Development of Pupil Personnel Services: An Operating Manual.** Massachusetts: North-eastern University, 1973.

Dayton, C.R., Helliwell, C.B., et al **Determining Program Goals And Objectives. Staff Training Module.** Mesa, Arizona: Mesa Public Schools, 1973.

Mesa Public Schools: **Toward Accountability: A Report on the Mesa Approach to Career Guidance, Counseling, and Placement.** Arizona: Author, 1973.

Pennsylvania Department of Education: **Guidelines for Self-Study of a School District Guidance Program.** Harrisburg, Pennsylvania Author, 1974.

South Dakota Department of Education. **Counselor's Workbook.** South Dakota: Guidance Services Division of the Elementary and Secondary Education, 1974.

Stufflebeame, D.L. **Evaluation as Enlightenment for Decision-Making.** Address delivered at the Working Conference on Assessment of Educational Outcomes, Sarasota, Florida: January, 1968.

PHILIPPINE JOURNAL OF
EDUCATIONAL MEASUREMENT