

COMPARISON AMONG TEACHERS IN
KHON KAEN, THAILAND TO DETERMINE
THEIR TESTING NEEDS

A Dissertation
for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
Sor-Wasna Pravalpruk
1974



3 1293 10019 1877



This is to certify that the

thesis entitled

COMPARISON AMONG TEACHERS IN KHON KAEN, THAILAND
TO DETERMINE THEIR TESTING NEEDS

presented by

Sor-Wasna Pravalpruk

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Education

William McArns

Major professor

Date Aug 8, 1974



MSU

OVERDUE FINES ARE 25¢ PER DAY
PER ITEM

Return to book drop to remove
this checkout from your record.

41

3:30

G

R

247

270

ABSTRACT

COMPARISON AMONG TEACHERS IN KHON KAEN, THAILAND TO DETERMINE THEIR TESTING NEEDS

By

Sor-Wasna Pravalpruk

In-service programs in educational measurement, with emphasis on test construction procedures, i.e., test planning and curriculum analysis, test item editing, item analysis, and reporting test scores, have been given to teachers in various parts of Thailand by the Ministry of Education and the Test Bureau at the College of Education, Prasarnmitr. It was the purpose of this study to find out which group of teachers in Khon Kaen province actually needs in-service training and in which areas of measurement the need is greatest. This study also yields some follow-up information on the effects of the earlier in-service programs. Two types of needs, actual and perceived needs, were investigated.

A Likert type questionnaire was constructed to measure the perceived needs. The items were a random sample of the tasks usually involved in the process of test construction. The respondent was asked to rate each task from 1 to 5, where 1 meant most needed and 5 not needed at all. The first edition of the questionnaire had eight items for each of the areas of in-service training for a total of thirty-two items. The actual needs were measured by a random sampling of test items measuring knowledge on educational measurement similar

to those items used in introductory measurement courses at Michigan State University. There were thirty-two items, eight each under the four areas of the in-service program. Both questionnaire and test were tried out and the reliabilities were found to be .87 and .41 respectively.

The revised editions of the questionnaire and test had twenty-eight items and twenty items respectively. They were sent together with the demographic data sheet and introduction letter to 400 teachers who were randomly selected from eight strata in Khon Kaen province. The stratification was based on the three variables: location of school university area or out of university area; levels of teacher education holders of teacher certificate and lower or holders of higher teacher certificate holders; and teacher experience less experience or more experience using eight years as the dividing line. Responses were received from teachers forming an equal sample size for each of the eight strata.

Responses were coded and data cards were produced. All canned programs and FORTRAN programs were run on the CDC 6500 at the Computer Center at Michigan State University.

The results of the descriptive information from the demographic data sheet indicated that the teacher-to-students ratio from the sample was lower than the ratio from the population (1:31 and 1:33 respectively). The correlation between total actual needs and total perceived needs was very low, i.e., .004. The correlation between the number of teachers in the respondent's school with perceived needs was low and negative at $-.069$ and with actual needs at $.160$.

The test of homogeneity of the variance-covariance matrix

was not significant for the four subscales of the actual needs but was significant for the perceived needs. The univariate repeated measures analysis was applied to the actual needs and the multivariate test to the perceived needs. The three way factorial design ($2 \times 2 \times 2$) with four repeated measures were analyzed by the FINN program. The design was crossed and balanced.

The analysis of the actual needs showed that the interaction among location of schools, level of teacher education, teaching experience, and the subject matter in measurement was significant. Descriptive follow-up showed that the less experienced teachers holding only the teacher certificate who worked within the university area and the more experienced teachers holding the higher teacher certificate who also worked within the university area had the most contrasting profiles. The four way interaction was explained by the significance of the location by teacher education by teacher experience interaction on the contrast variable of Test Planning and Reporting Test Scores. The three-way interaction was not significant when Test Planning was analyzed individually, but the interaction was significant for Reporting Test Scores alone. In the presence of the four-way interaction, the subject matter by the level of teacher education by experience interaction, the subject matter by experience interaction, the subject matter by location interaction, the subject matter main effect, the location by experience interaction and the teacher education main effects were significant at $\alpha = .05$. All interactions were presented in graph form.

On perceived needs, only three null hypotheses were rejected. The subject matter by location interaction was studied along with

the subject matter main effect. For teachers in the university area, the more experienced teachers had perceived needs ranging from highest to lowest in the following order: Item Analysis; Reporting Test Scores; Test Planning; and Item Editing. But for those outside the university area, the perceived needs on Item Editing was second, followed by Reporting Test Scores and Test Planning. In general, the Item Analysis had the highest perceived needs. The other null hypothesis rejected was the location of schools by teaching experience interaction. The interaction was disordinal which could be explained by the observation that less experienced teachers within the university area had lower perceived needs than the more experienced teachers within the university area but less experienced teachers outside the university area had higher perceived needs than more experienced teachers outside the university area. The difference between perceived needs of all (both inside and outside the university area) of the less experienced teachers was smaller than all of the more experienced teachers.

COMPARISON AMONG TEACHERS IN KHON KAEN, THAILAND
TO DETERMINE THEIR TESTING NEEDS

By

Sor-Wasna Pravalpruk

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services, and Educational Psychology

1974

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to my Advisor and Chairman of my study, Professor William A. Mehrens for his assistance and suggestions throughout my program of study and during the preparation of this dissertation. To all the members of my doctoral committee, I am grateful for their contribution of encouragement and insights. To Dr. Archibald B. Shaw, who was kind enough to spend many hours in correcting my English. To Drs. Maryellen McSweeney and Willard Warrington for their help in the development and improvement of the study.

Appreciation is also conveyed to Nitra Somswasdi, my sister, for her help in distributing and collecting the questionnaire. To Dr. Kowit Pravalpruk, my husband, for his help in analyzing the data. His constant support and encouraging me to persist with my study were very much appreciated.

I acknowledge with appreciation the support of the Thai Government which allowed me to pursue my doctorate study. Special thanks are also extended to Terry Nafisi-Movaghar who typed with patience this dissertation.

Finally, for their encouraging, I wish to extend my gratitude to my parents, my brothers and sisters, and to Puttaporn, my daughter.

TABLE OF CONTENTS

<u>Chapter</u>		<u>Page</u>
I	The Problem	1
	Introduction	1
	Need for the Study	2
	Thai Educational System	3
	Purpose of the Study	7
	Hypotheses	8
	Definition of Terms	9
	Overview	10
II	Review of the Literature	11
	The Importance of Classroom Testing	12
	The Qualifications of Teachers as Evaluators	15
	In-Service Programs in Testing and Evaluation	18
	Summary	24
III	Procedure	27
	Population	27
	Sampling Procedure	28
	Instrument	30
	Data Collection	34
	Design	35
	Analysis	38
	Summary	39

<u>Chapter</u>		<u>Page</u>
IV	Analyses and Results	40
	Descriptive Data	41
	Homogeneity of Variance-Covariance Matrices	45
	Repeated Measure Analysis on Actual Needs	46
	Repeated Measure Analysis on Perceived Needs	60
	Additional Analysis	67
	Summary	73
V	Summary and Conclusions	79
	Summary and Discussion	79
	Conclusions and Implications	84
	Recommendations for Further Study	88
Appendix A	The Instrument (First Edition)	92
Appendix B	The Instrument (Final Edition)	103
Appendix C	Pairwise Profile Analysis on Actual Needs	124
Appendix D	Multivariate Analysis of Variance on Four Subscales on Actual Needs	126
Bibliography		128

LIST OF TABLES

<u>Table</u>		<u>Page</u>
3.1	Total Number of Teachers in Khon Kaen Classified by Location of School, Class of Teacher Certificate and Years of Teaching Experience	29
3.2	Design of the Three Independent Variables: Location of Schools, Levels of Teacher Education, and Teaching Experience	36
4.1	Averages and Standard Deviations of the Number of Teachers and Number of Students in Sampled Schools Classified According to Eight Groups of Teachers	42
4.2	Inter-Correlations among Number of Teachers, Number of Students, Perceived Needs, and Actual Needs	43
4.3	Variance-Covariance and Correlation Matrices of Four Test Scores	47
4.4	Variance-Covariance and Correlation Matrices of Four Perceived Needs Scores	48
4.5	Means and Standard Deviations of Four Test Scores (Actual Needs)	49
4.6	Univariate Repeated Measured Analysis on Test Scores	51
4.7	Means and Standard Deviations of Four Perceived Needs Scores	61

<u>Table</u>	<u>Page</u>
4.8	Univariate and Multivariate Repeated Measured Analysis on Perceived Needs 63
4.9	Comparison on Perceived and Actual Needs between Teachers Who Attended and Did Not Attend the In-Service Program 68
4.10	Comparison on Perceived and Actual Needs between Teachers Who were Favorable and Not Favorable to the National Testing Policy 69
4.11	Comparison on Perceived and Actual Needs between Male and Female Teachers 70
4.12	Comparison on Perceived and Actual Needs among Teachers Who Taught Different Grade Levels . . . 71

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
I Graph Presentation of Cell Means for Subject Matter, Teacher Education, and Experience (Actual Needs) . . .	55
II Graph Presentation of Cell Means for Subject Matter and Teaching Experience (Actual Needs)	57
III Graph Presentation of Cell Means for Subject Matter and Location of Schools (Actual Needs)	58
IV Graph Presentation of Cell Means for Location of Schools and Teaching Experience (Actual Needs)	59
V Graph Presentation of Cell Means for Location of Schools and Subject Matter (Perceived Needs)	64
VI Graph Presentation of Cell Means for Location of Schools and Experience of Teachers (Perceived Needs)	65

CHAPTER I

THE PROBLEM

Introduction

Man has always been concerned with measurement and evaluation. From the earliest beginnings of society, men have measured the abilities of other men and have recognized the existence of differences in the abilities possessed by different individuals. As human society has grown in size and complexity the recognition of individual differences has increasingly been reflected in the structure of societies.

In school, the teacher employs a central role in the testing and evaluation process. The teacher must make judgments, measure, appraise and report. To do this he must use tests, examinations, inventories and other types of measuring instruments and devices. He must know how to construct such instruments for his own purposes, and to use and interpret the results of others. He must know something about the analysis and interpretation of scores on tests.

Classroom testing not only provides an expedient method for grading students, it also gives the student a far more valid indication of what the teacher considers important than do statements made in the textbook or by the teacher. Because it provides the basis for grades, the test indicates unmistakably to the student what should be learned or what is expected. Tests can also provide the teacher with needed information on precisely what students are learning and what they are not learning.

Need for the Study

Tests have become an integral part of our academic life. Teachers evaluate students from the time they are in kindergarten until they leave high school or college. Although decisions are often made on the basis of a student's scores, little, if any, thought is given to the qualifications or skill of the teacher as an evaluator. It is essential that we recognize the value that test results may play in the life of the student.

Teachers occupy a central role in the testing and evaluation process. Teacher-made tests are frequently the major basis for evaluating the progress of students. They may be better than good commercially prepared achievement tests, in the sense that they may be more relevant to a teacher's particular objectives and pupils. Unfortunately, many teacher-made tests do not serve their function optimally because very often teachers have not been given adequate preparation in this area of competence. Two surveys done about 20 years ago (Noll, 1955; Allen, 1956) were in substantial agreement in showing that an introductory course in measurement is not generally required by State Departments of Education for a teaching certificate. Most institutions preparing teachers offer an introductory course in educational measurement but comparatively few require it for a recommendation for a teaching certificate. More recent studies done by Goslin (1967) and Roeder (1972) lead to the same conclusion, that the problem of preparing teachers in measurement and evaluation is real, clear, and substantial. Many teachers do not seem to be adequately prepared in this respect, and apparently they recognize their inadequacy.

If we accept testing and the use of test results as a part

of the teachers' role, efforts need to be made to make sure that this part of their role is not carried out in a context of naivety about tests and what they measure. The role of teachers in testing is too important to be left to chance.

In many systems, in-service courses are offered to teachers from time to time. A varying number of sessions is held according to the other demands being made on teachers. This type of program might be useful to help school personnel become more familiar with the theory and principles of measurement and evaluation.

Thai Educational System

The educational system in Thailand is quite different from the United States. Historically, the Thai educational system has had a strong connection with Buddhism, the main religion in Thailand. The Wat (Buddhist temple) was the only place available for youngsters to be educated formally and the Buddhist monks were the only teachers. The subject matter was Buddhist philosophy, history, and theology. Vocational education was either taught by a skillful person in that particular job or it was passed along within the family. Only a comparatively few youngsters and only boys, were given an opportunity to be educated in schools. Education for most was just on-the-job training or part of the living in the family occupation. Most of the boys learned to do as their fathers did, and the girls learned only the skills of house-work and home crafts. Therefore, testing for most was in the form of tests of performance and evaluation was just "it is the time for a birdie to fly alone."

In 1871 the first school was established in the old palace

in Bangkok, the capital city of Thailand. King Rama V (Chulalongkorn) supported the school, and formal education became the royal family and royal cousins' affair. Going to school was aimed towards becoming a royal officer in one of the civil services. Learning was to read, to write, and to have the social graces needed to get along in the King's services. Schooling proved to be such a success and so useful to the development of the kingdom that the King proclaimed that schools be established throughout the country within a few years. By 1887 there was at least one school in each province. The curriculum was basically the three R's. A British-type school was implemented. From the beginning and down to today, youngsters go to school three quarters of the year. There is no school during the summer. There are tests at the end of each quarter but the final examination at the end of the school year is the basis for promotion to the next grade. Early examinations were either in the form of class recital and individual recital or in essay form when the recital examination was too time consuming. When the Ministry of Education was founded in 1892, a nation-wide system of examinations was put into practice. For a long time, all tests were in essay form.

In 1954 the objective test item was introduced. The initial effort was credited to M.L. Tuy Chumsai. With very little knowledge and understanding of the principles involved, teachers began to use objective tests. The lack of knowledge caused the production of bad objective test items. The first textbook on educational measurement in the Thai language was written by Chaval Paeratakul and published in 1963. Two years later, a nation-wide inservice program in educational measurement was proposed and carried

out by Chaval Paeratakul and the staff of the Test Bureau in the College of Education at Prasarnmit in an attempt to improve the quality of the teacher-made tests. During this period, the Test Bureau constructed four standardized tests for grades 4 and 7, two in the subject Thai Language and two in Arithmetic. The tests and their norms were published in 1971.

In 1965 in-service workshops were scheduled for 5 days, and covered these topics: purposes of measurement; curriculum analysis as a blueprint for test construction; types of test items; item analysis; scores and norms; and reporting the test result to parents and authorities. The morning sessions included lectures by specialists from the College of Education and the Department of Teacher Training. The afternoon sessions were practicums. Teachers were assigned to one of the small groups working on blueprints, writing test items, performing item analysis, and reporting standard scores. In the past five years, the training has emphasized test construction and writing test items according to Bloom's Taxonomy.

The quality of the test is critical because the evaluation that follows can slow down a student's progress through school by one year. If a student fails the final examination at the end of the school year he has to be in the same grade for another year with some unfavorable attitude and the label "repeater." Since education in Thailand is in effect a selective system, less than 7 percent of the first graders will finish early secondary school (grade 10). Entrance examinations are used to select students who want to go on to upper secondary school (grades 11 and 12). Another entrance examination at the end of grade 12 is used for admission to

universities. Good tests are needed in the selection process so that educational wastage is minimum.

In school, a student is mostly recognized and rewarded by teachers, parents, friends, and community on his cognitive abilities. Those cognitive abilities are measured by teacher-made tests. Again the quality of the test is crucial. Tremendous pressure is on a student whenever he is going to take a test. Sometimes it means his entire future depends upon one or two hours of his performance on the test. What will happen if one is not feeling well when the time comes to take the test?

In 1962 the Department of Teacher Training implemented a new curriculum in both two and four year programs of teacher education. The new curriculum added one course in educational measurement and evaluation in the four year program. This course covered how to construct an objective test from planning to interpretation of the test score. It was 1966 when the first graduates from the new curriculum were assigned to schools. It was hoped that these new teachers would help other teachers in the same school learn how to construct objective tests. However seniority is a very real barrier in communication between old teachers and new teachers. If the seniority barrier assumption is true, there will be a difference in the new teachers' needs and the older teachers' needs for an in-service program in testing. It is interesting to investigate whether the older teachers need more testing services.

Because of the centralized system of education in Thailand, information and policy are diffused from the Ministry of Education to the Provincial educational officers and then to the teachers.

Within one province the central amphur will receive the information at first-hand. Other amphurs in the province will get them later depending on how far they are from the central amphur. In Khon Kaen, there is a university which is located in the central amphur. These two facts lead to the conclusion that teachers in the central amphur should have had more help and can have acquired knowledge about educational measurement easier than those in other amphurs. Therefore, they should have less need for in-service courses than teachers who are not in the university area.

The level of the teachers' certificate furnishes another way to classify teacher groups. The holders of higher certificates took more courses in their four-year program and should have more knowledge about educational measurement. Teachers who hold the certificate gained in a two-year program, it is hypothesized, may have more need for help than do the holders of higher teacher certificates.

As a result of this study, one should be able to conclude a) which group of teachers in Khon Kaen Province has the most urgent need, and b) what area of service is most needed. It is hypothesized that teachers who have Teacher Certificates or lower, who are not in the university area and who are older teachers (teach more than 8 years), have the highest need, and that writing the objective items will be the most needed area of service.

Purpose of the Study

The purpose of this study is to identify the areas of instruction in measurement which are needed most by the teachers in

the Thailand province of Khon Kaen, and also to compare teachers who are grouped according to the amount of their experience in teaching, to the location of their schools, and to the kind of teaching certificate they hold. Actual and perceived needs in four fundamental areas will be measured: planning test construction or curriculum analysis, test item editing, item analysis, and reporting test scores.

Hypotheses

The major hypotheses to be examined are as follows:

1. Teachers who are in the university area may have fewer actual and perceived needs than teachers who are farther from the university in Khon Kaen Province.
2. Teachers who hold higher teacher certificates may have less actual and perceived needs than teachers who hold the teacher certificate in Khon Kaen Province.
3. Older teachers (more teaching experience) are expected to have more actual and perceived needs than newer teachers in Khon Kaen Province.
4. No difference is expected in actual and perceived needs among the four subject matters covered in the study.

Besides these four hypotheses about main effects, all interactions will be tested. These interaction effects are determined in two parts: the interactions between the three variables of location, level of teacher education and level of experience; and the interactions between the three variables and the repeated measure variable, the four subject matter areas covered in this study.

Definition of Terms

Actual need is defined as a lack of knowledge in a sub-area of tests and measurement as is disclosed by responses to groups of related test items that are designed to test desirable knowledge of measurement and evaluation.

Perceived need is defined as the feeling of lacking desirable knowledge in tests and measurement as measured by a Likert Type format questionnaire about tasks in measurement and evaluation.

Amphur is a political subdivision of a province. There are 13 amphurs in Khon Kaen Province.

University Area is the Central Amphur of Khon Kaen Province. The university is Khon Kaen University.

Far From University Area describes all amphurs in Khon Kaen Province except the Central Amphur. It is also called outside university area.

Certificate teacher is a teacher who graduated from a two-year teacher training curriculum, or its equivalent.

Higher certificate teacher is a teacher who graduated from a four-year teacher training curriculum, or its equivalent.

New teacher is a teacher who has been teaching for less than eight academic years.

Old teacher is a teacher who has been teaching for eight or more academic years.

Overview

The study is organized into five chapters, followed by appendices.

In Chapter II, the literature relevant to the general problem and related areas is reviewed. The design of the study, the pilot study, the instrumentation, and the methods of analysis are discussed in Chapter III. The research data and results of the study are presented in Chapter IV. The final chapter contains a summary of the study, the conclusions, a discussion of the findings and implications for future research.

CHAPTER II

REVIEW OF THE LITERATURE

The intent of this chapter is to review studies that are related to the problem as described in Chapter I. The review is divided into three sections: the importance of classroom testing; the qualifications of teachers as evaluators; and in-service programs in testing and evaluation.

The Importance of Classroom Testing

It is generally agreed that evaluation of pupil progress is a major aspect of the teacher's job. A good picture of where the pupil is and of how he is progressing is fundamental both to effective teaching and to effective learning.

Classroom testing may be used in the form of quizzes, midterm and final examinations, and as general guideposts for teaching. Tests may be given at various times during the school year in order to stimulate learning through review and to determine status and progress in learning. As Mehrens and Lehmann (1973, p. 167) state: "Teacher-made tests are frequently the major basis for evaluating the students' progress in school."

Teacher-made tests, by their definition, are constructed by a classroom teacher for use in his particular classes under conditions of his choosing. Classroom achievement tests, if properly prepared, are tailor-made for a certain class of pupils taught in a certain manner. They can be related to the kind of educational objectives

involved in the class and also to the relative emphasis each received. They also can be shaped to fit the particular characteristics of the pupils who compose the class. The teacher is in the best position to know the exact nature of the characteristics of his class. Further, since the appropriateness of any achievement test item is determined by the educational objectives of the class, the teacher is in an excellent position to build test items that measure suitable areas of the pupils' achievements and that are of suitable difficulty.

Commercially prepared achievement tests may be used to obtain some of the information needed by teachers, and they may even be used to motivate students. But it is unusual for standardized tests to be administered more than once a year. Teacher-made tests are better in the sense that they are more relevant to a teacher's particular objectives and to his class. As Mehrens and Lehmann (1973, p. 202) say, "Classroom tests, despite some of their limitations, will never be replaced because they (a) tend to be more relevant, (b) can be tailored to fit a teacher's particular instructional objectives, and (c) can be adapted better to fit the needs and abilities of the students than can commercially published tests."

Also Ebel (1962, p. 24) states:

Even if good ready-made tests were generally available, a case could still be made for teacher-prepared tests; the chief reason being that the process of test development can help the teacher define his objectives. This process can result in tests that are more highly relevant than any external tests are likely to be. It can make the process of measuring educational achievement an integral part of the whole process of instruction as it should be.

Schwartz and Tiedeman (1957, p. 110) say that teacher-made tests are valuable because they

1. Assess the extent and degree of student progress with reference to specific classroom activities.

2. Motivate students.

3. Permit the teacher to ascertain individual strengths and weaknesses while the pupil is studying a particular subject-matter area.

4. Provide information for reporting purposes.

5. Provide immediate feedback for the teacher, who can then decide whether a particular unit or concept needs reteaching and/or whether individual pupils are in need for remedial instruction.

6. Provide for continuous evaluation.

A good classroom test must be valid, reliable, and objective. It must possess good discriminative power, and it must be characterized by questions with a range of difficulty suitable to the abilities of all the students in the class. Bell (1972, p. 8) says: "A teacher who gives tests and uses them in the classroom without ever knowing whether and to what degree such qualities are present is flying blind."

Teacher-constructed tests should be designed to measure areas of strengths and weaknesses, and thus enable the teacher to plan for a better program of organization, preparation, or presentation, for the purpose of reteaching. Furthermore, Thorndike and Hagen (1969, pp. 30-33) suggest that functions of teacher-made tests are:

- a) Motivation.

"Tests that are well constructed and effectively used can motivate students to develop good study habits, to correct errors, and to direct their activities toward the achievement of desired goals." (p. 31)

- b) Diagnosis and instruction.

"The items on which an individual fails or on which many members of a class group fail can serve to identify points needing further study." (p. 31)

c) Defining teaching objectives.

"We may know a teacher by the tests he makes. They tell what he is truly valuing in his pupils, even though he himself does not know it, and they influence profoundly what his students will learn." (p. 32)

d) Differentiation and certification of pupils.

"The teacher inevitably has a responsibility for certifying pupils' accomplishments to higher levels of the educational enterprise and to the world outside the school." (p. 33)

The authorities seem to agree that in constructing a test, the teacher must be concerned with the same factors as are the author and publisher of a standardized achievement test—validity, reliability, and objectivity.

Evaluation of pupil achievement, in their view, is one of the teacher's important responsibilities. It is important that the teacher's device be well thought out and well made. For any type of written test, it is desirable to have a definite plan in advance of preparing the test items. The development of such a plan requires an analysis of the outcomes one is trying to achieve in the teaching of a particular course or unit and of the significant segments of content through which those objectives are to be realized. In addition, the plan should include the allocation of test items among the content areas and objectives, the types of items to be used, the total numbers of items in the test, and specifications for the spread of item difficulties.

In Thailand, more emphasis is on the written examinations in which teachers have to depend on the tests they have made individually. Evaluation at the end of the school year is based almost entirely on the teacher-made tests (Paeratakul, 1963, in Thai edition). Only

the performance during the five days of the last week of the school year is the basis for determining which students will go to the next grade or repeat the same grade in the next school year. So there is much more stress on teacher-made tests in Thailand than in the United States.

The Qualifications of Teachers As Evaluators

Teachers occupy a central role in the testing and evaluation process. Teacher-made tests are frequently the major basis for evaluating the progress of students. Many teacher-made tests do not serve their function optimally because of time pressures on teachers and, more important, because test-writing is not generally a part of the professional training of teachers (Grobman, 1971).

In 1955 Victor Noll reported a study of requirements in measurement by state departments of education for teaching certificates and by 80 institutions preparing teachers in all parts of the country. The sampling included large publicly supported institutions, large privately supported institutions, teachers colleges, and liberal art colleges. A similar survey was reported in 1956 by Allen. The results in her study were based on questionnaire returns from 288 higher institutions with programs for the preparation of teachers, an 88 percent return. The findings of the two surveys are in substantial agreement. They showed that an introductory course in measurement was not generally required by state departments of education for a teaching certificate. Most institutions preparing teachers, they found, offer an introductory course in educational measurement but comparatively few require it for a recommendation for

a teaching certificate.

Another study done under the Russell Sage Foundation in 1967 by Goslin states that

...with respect to formal training in test and measurement techniques (courses taken in college, attendance at clinics on standardized testing, and the like), the data indicate that less than 40 percent of all teachers have had more than minimal exposure (one course) to such training, and that a sizable proportion of teachers have never had a course in measurement techniques or attended a clinic at which testing was discussed. Data also indicate that teacher involvement in testing is both high and of potentially great influence on the educational system and the children within it. If we accept testing and the use of test results as a part of the teacher role, efforts need to be made to make sure that this part of their role is not carried out by chance. (1967, p. 127)

A more recent study was conducted by Roeder (1972) who surveyed the qualifications or skills of the teacher as an evaluator. The 940 elementary teacher-training institutions located in every state and the District of Columbia, were mailed a one-page questionnaire. A total of 916 responses (97.4 percent) were received. The data indicate that 57.7 percent of the institutions which were surveyed, or 496 institutions, did not require their prospective elementary teachers to complete a course in evaluation; 12.1 percent (104) required nothing more than a one or two semester hour course; 17.8 percent (158) required a three semester hour course and only 1.4 percent (12) required four or more semester hours or course work in evaluation. Roeder concludes that the majority of teachers receive only a minimal exposure to the complex world of evaluation. That is, most of today's elementary teachers are not prepared to make or use tests.

Grobman (1967, 1971) sees testing as an integral part of teaching. She found that many teachers testified that the improvement of their test building skills has resulted in the improvement of their teaching, since it has made them more sensitive to their teaching aims and the implementation of these. She claims that few teachers have had an opportunity to acquire a sufficient background in evaluation and testing to enable them to measure adequately the skills and understanding they wish to encourage in their students.

Ebel (1961, p. 68) states that a teacher who is competent in educational measurement should:

- (1) Know the educational uses as well as the limitation, of educational tests.
- (2) Know the criteria by which the quality of a test should be judged and how to secure evidence relating to these criteria.
- (3) Know how to plan a test and write the test questions to be included in it.
- (4) Know how to select a standardized test that will be effective in a particular situation.
- (5) Know how to administer a test properly, efficiently, and fairly.
- (6) Know how to interpret test scores correctly and fully, but with recognition of their limitations.

Further, in the 62nd Yearbook of the National Society for the Study of Education (1963) Hagen and Lindberg express a belief that classroom teachers: (1) must have a basic understanding of the evaluative process, including an understanding of the basic concepts of reliability, standard error of measurement, and validity as essential; (2) should know the general types of tests available and be aware of their uses, strengths, and limitations; and (3) should be able to

combine available test data with other records and to interpret them as they relate to individual children in their classes.

Chaval Paeratakul (1963), the first person from Thailand who graduated in Tests and Measurements in the United States, along with other professionals in this area agreed that there was an urgent need for teachers in Thailand in tests and measurements. Other evidence to support this idea was the change in the curricula of teacher training education programs in 1962 and the nation-wide in-service program from 1965-1966. Teachers in Thailand have just begun to know what the objective tests are all about. They definitely need more help in getting acquainted with these subject matters.

In-Service Programs in Testing and Evaluation

From studies reviewed in the last section, it is clear that the problem of competence of teachers in educational measurement and evaluation is real and substantial. They lead to the conclusion that many teachers do not seem to be adequately prepared in this respect and apparently they recognize their inadequacy.

It is of the utmost importance to educational progress that the competence of teachers to measure educational attainments be improved. Ebel (1961) says that far more harm, perhaps ten times as much harm, is currently being done to student learning as a result of the shortcomings of the classroom tests by which a student's educational efforts are largely stimulated, directed, and evaluated, than is being done by all the faults of external testing programs.

✓ He suggests in-service programs as a means for improving the competence of teachers in measurement on the basis of experience with a variety

of in-service programs intended to help teachers solve measurement problems. These programs, variously referred to as conferences, seminars, institutes, or workshops, have ranged from an afternoon lecture to a three-day preschool program with several follow-up meetings later in the year. Some of these programs were sponsored by a single school system and involved the teachers of that school in all subject areas and at all levels. Others reflected the interest of a single professional group, such as engineers or nurses.

Goslin (1967, p. 129) suggests the publication of a short booklet designed explicitly to deal with the problems faced by teachers in their role as testers as a supplement to a "more formal training in testing" to prospective teachers.

Hagen and Lindberg (1963) claim that each school system must take responsibility for providing orientation and in-service training in testing because it is unrealistic to assume that teachers can learn all that they need to know about testing during their pre-service courses.

There was a panel discussion about the improvement of the competence of classroom teachers in measurement and evaluation held by the National Council on Measurements Used in Education, at the 1959 annual meeting. The discussion centered on the questions of why in-service training is a problem, and what can be done to improve in-service training at the local level.

As an example one local in-service program was presented in specific steps:

- 1) Preparation of local bulletins. A series of bulletins, supplementing the published materials concerning the tests in use in the county, have been written tying in the testing program with the local program of instruction.
- 2) Use of school test coordinators. These test coordinators met regularly, especially before and after a scheduled testing program, to discuss problems involved in administering, scoring, and interpreting the test results.
- 3) Extension courses.
- 4) Faculty workshops. A considerable number of faculty workshops, varying in length from one to four or five sessions, have been held.
- 5) Development of community interest in the measurement program, through a judicious use of local newspaper publicity, talks to parent-teacher associations.
- 6) Use of local norms.
- 7) Provision of adequate physical facilities.
- 8) Use of demonstrations, lectures, etc., in the district's In-Service Training Center.
- 9) TV Workshop on testing.

(Durost, pp. 32-33)

Rather than the in-service program, McSwain (1959) focused on the role of the college faculty in the teacher education program. He said that during the period of student teaching each student-teacher should learn, from teacher and faculty advisor, about these areas: (1) a functional concept of evaluation and measurement in creative teaching and purposeful self-education; (2) techniques used in evaluating the behavioral responses of pupils; and (3) criteria to be applied in selecting and using educational measures to improve teacher diagnosis and effective procedures in preparing and interpreting cumulative child-accounting records. And, he added,

the degree of competence in applied measurement developed by each student during this professional internship depends basically upon the leadership of the faculty counselor.

In the same panel discussion, Charles Stevenson (1959) talked about measurement in the teaching-learning situation. He stated that the in-service program should cover: (1) common understanding of what is to be measured among teachers, supervisors, coordinators, and superintendents; (2) the type of evaluative instruments which will be adequate in a particular situation; (3) clarification of how the results of measurement can be used to improve the teaching-learning situation; and (4) methods and procedures to be employed in improving these techniques of measurement.

The last presentation in this panel, by Margaret Stevenson, was a discussion of the role of the classroom teacher in school testing programs. She suggested that after the school system has formulated its philosophy on testing; after representatives of all levels of school personnel have cooperated in developing the program; after the testing program has been given a definite place in the school curriculum--then the major hurdles have been passed, and the school system is ready to address itself to the more direct problem of improving classroom teacher competence in measurement and evaluation. That is, improving classroom teacher competence in measurement and evaluation must not be treated as an isolated question. It must be viewed in context, against the background of the numerous problems involved in establishing an adequate, democratically structured testing program in the school curriculum.

The importance of in-service education for all educational

personnel is recognized throughout the literature of the teaching profession (Ebel, Ed., 1969, p. 645). The rapid expansion of knowledge, which has been reported extensively over the past several years, and its effect on changing methods and in developing technology utilized in the classroom are major factors in making programs of in-service education necessary.

A statewide research study in Tennessee about in-service education was conducted by Brimm and Tollett (1974). The conclusions of the study can be summarized as the following:

- (1) The primary purpose of in-service education is to upgrade the teacher's classroom performance.
- (2) If continuing professional growth is to be taken seriously, administrators and teachers must pool their knowledge and resources and seek to make in-service programs more responsive to the needs and interests of practicing classroom teachers.

The two chief weaknesses of the in-service program as Ebel (1961) sees them are: (1) they are too brief. While an hour or two a year spent in considering measurement problems under the guidance of a specialist is far better than nothing at all, it is unreasonable to suppose that satisfying, enduring progress in solving the manifold problems of educational measurement, or in developing the requisite knowledge, understanding, and skills, can be made in so short time; and (2) they involve too much talking and too little doing. For the cultivation of a practical art like educational measurement, sound pedagogy requires a mingling of theory and practice. Ebel presents the ideal program of in-service training for improving the competence of teachers in measurement as the following:

Suppose that a school administrator and his staff have decided to focus attention for a year on the improvement of classroom testing. Suppose they engage a specialist in educational testing to meet with them five times during the year, at intervals of six weeks or so, for a day or two. Participation in the initial program might well be limited to five, six, or seven groups of four to six teachers each.

The goal of each group would be to make, to use and to analyze a quality test in a subject which all members of the particular group were teaching. Examples of subject area in which these tests might be developed are: fourth grade mathematics; sixth grade geography; eighth grade English; or high school history; chemistry; or economics.

The first meeting of each participating group would be devoted to a description of the entire project, with special consideration of the first step - the preparation of specifications for the test to be developed. Sample specifications would be presented for study and analysis. Between the first and second sessions each teacher group would work out the specifications for its test. These could be reviewed at the second meeting, and work on item writing would be launched. The third meeting could be devoted to item review and test assembly, the fourth to test administration and analysis, and the fifth to a review of the test developed and of the entire project as a learning experience. (Ebel, 1961, pp. 70-71)

A program like this, Ebel says, would produce not only a handful of excellent tests but also a sizable group of teachers whose competence in measurement was vastly improved and highly respectable.

One of the most effective in-service programs in educational measurement in Thailand was carried out by Chaval Paeratakul in 1965-66.

✓ The purpose of the program was to improve the quality of the teacher-made tests. It was a five-day workshop. The material covered included: purposes of measurement; curriculum analysis as a blueprint for test construction; types of test items; item analysis; scores and norms; and reporting the test result to parents and authorities. The morning sessions were lectures by authorities.

The afternoon sessions were practicums. Teachers were assigned to one of the small groups working on blueprints, writing test items, performing item analysis, and reporting standard scores. The training had emphasized test construction and writing test items according to Bloom's Taxonomy.

Summary

In most of the literature, educational measurement and evaluation are viewed as essential and constructive (Ebel, 1972; Schultz, 1971; Sharp, 1966; Jessel, 1972; and Mehrens, 1967). Occasionally some suggest that education could go on perfectly well, perhaps much better than it has in the past, if tests and testing were abolished. This may imply that the tests these critics have seen or have used themselves are not good. No doubt some bad tests have actually been worse than no tests at all.

Test scores are powerful educational tools, but they are only tools. They will not give direct, complete answers to the practical educational questions. They will not point unequivocally to a specific course of action in a given set of circumstances. They may indicate that something needs to be done. They may provide data that will help in deciding what to do. But they will not make the decision.

Several writers (Mehrens, 1967; Trump, 1967; Holt, 1971; McMorris et al, 1972; and Dyer, 1962) suggest that the limitations and underlying principles of tests must be recognized if one wants to use tests as good measuring tools in education.

The qualifications of teachers as evaluators are widely

reported as less than satisfactory for most teachers. Studies on competence of teachers in measurement and evaluation conclude that most teachers do not have adequate knowledge in this area (Goodman, 1971; Noll, 1955; Allen, 1956; Goslin, 1967; and Roeder, 1972). These studies also suggest that more publication and service programs need to be operated to improve the situation.

The authorities conclude that in-service or workshop programs are needed to improve the competence of teachers in measurement. The purpose of the program is primarily to help teachers solve measurement problems. Much time, effort, and money needs to be spent to initiate or improve programs of services and assistance to teachers. The justification is that this is an important method of raising the standards of education.

These help programs may be as little as a short booklet (Goslin, 1967), or they may be conferences, seminars, or workshops which range from one afternoon to a five-session workshop (Huddleston, 1959; Ebel, 1961). The contents to be covered are the basic knowledge in educational measurement and evaluation that teachers should know, such as the limitations of educational tests, the properties of tests, ways of planning and constructing tests to use for one's own purposes, test administering, and how to interpret test results.

Teacher-made tests in Thailand are much more important to effective teaching and to effective learning than in the United States since standardized tests are not available yet. And, also, more stress is put on the final examination than in the United States. The teacher-made final examination has at least seventy-five percent weight on evaluating "pass" or "fail" for the

students. The qualifications of Thai teachers as evaluators are much worse than in the United States. Also, the need for improving the competence of teachers in Thailand is very much greater than in the United States.

CHAPTER III

PROCEDURE

This study can be classified as a comparative study. Questionnaires, sent through the mail or delivered personally, were used to collect the data. This chapter provides a description of the population, sampling procedure, instrument, data collection, and plan for data analysis. Findings in the study are presented in Chapter IV and conclusions are given in Chapter V.

Population

Geographically, Khon Kaen province is in the northeastern part of Thailand, about 200 miles from Bangkok by the "Friendship" (Mitraparp) highway. In area, the province is 14,076 square-kilometers. The province is divided into thirteen amphurs and these thirteen amphurs have 116 tambons (subdivisions of amphur). The population is 1,136,939 people, the majority of whom are farmers.

Educationally, there are 1047 schools, one technical institute, and one university. Of the 1047 schools, thirty-five offer a high school program either alone or in combination with the elementary and secondary levels in the same school. Forty-three of 1047 schools are schools for adults, which have both elementary and secondary levels.

The population of interest is teachers in these 1047 schools. There are 5,913 teachers and 198,246 students. The average ratio of teacher to student is 1:33. The breakdown of teachers into

eight groups is presented in Table 3.1. The largest group (1534) is composed of teachers who are not in the central amphur, who have been teaching more than eight years and who hold the teacher certificate or lower. It is estimated that more than ninety percent of these teachers are in elementary schools or teach in elementary grades. Most of the big schools (more than 1000 students) are in the university area, so the number of teachers in the central amphur is the largest among the thirteen amphurs. The smallest group (210) is composed of teachers who are in the university area, who hold at least the higher teacher certificate, and who have been teaching more than eight years.

Of the 5,913 teachers, sixteen hold master's degrees and 349 have at least the baccalaureate degree. Six percent of all teachers (365) hold at least baccalaureate degrees and most of them are teaching in the university area. Eighty-nine percent of the teachers (5249) teach at the elementary level (grades 1 to 7) and seventy-one percent (4212) in lower elementary level (grades 1 to 4). Of the totals, 844 teachers are in fifty-eight private schools, mostly in the university area. All others are in public schools, either in provincial municipal sections or under the Departments of Elementary and Secondary Education.

Sampling Procedure

A list of all the teachers in Khon Kaen province was provided by the Ministry of Education. Along with the names, the location of schools, ages of teachers, years of teaching experience, and levels of teacher education were given. The teachers were classified into eight groups, according to location of schools,

TABLE 3.1

TOTAL NUMBER OF TEACHERS IN KHON KAEN
CLASSIFIED BY LOCATION OF SCHOOL, CLASS OF TEACHER CERTIFICATE
AND YEARS OF TEACHING EXPERIENCE

Location	Teacher Certificate or Lower		Higher Certificate		Total
	Less Than 8 Yrs. Experience	8 Yrs. or More Experience	Less Than 8 Yrs. Experience	8 Yrs. or More Experience	
In the University Area	749	589	335	210	1883
Not in the University Area	1203	1534	748	545	4030
Total	4075		1838		5913

years of teaching, and levels of teacher education. A tentative number was assigned to every teacher for the purpose of random selection. A table of random numbers was used. A systematic random sampling procedure was used to get a sample of 40 teachers in each of the 8 groups for the total of 320 teachers.

Because of the difficulty of transportation and uncertainty of the mail, it was expected that the percentage of return would be lower for the teachers who are outside than those inside the university area. Therefore, 80 more teachers were randomly selected, 20 from each of the 4 groups outside the university area. The first 40 returns of each of these 4 groups was used in the analysis. If less than 40 were returned, in any of the groups, additional numbers of teachers were selected and personal delivery questionnaires were distributed. However, only a very few additions were necessary. Since most teachers in Thailand are government officers and their jobs are alike because of the very strong centralized educational system, the biasedness contributed by substitution of the sampling is minimum according to the homogeneity of the population.

No problem of data collection was expected from those in the university area because most of the teachers received their questionnaires by personal delivery. Since the agent collecting the questionnaires lived in the university area, the follow-up was done weekly. Because of the personal relationship and with some help from the Head Masters, a hundred percent return was received in the University area.

Instrument

The instrument used was composed of two parts:

- 1) a questionnaire on perceived needs using a Likert type answer format.

- 2) test items about basic knowledge in measurement and evaluation.

There were four subject matters covered in the test and in the questionnaire. Each part was composed of seven items for the perceived needs and five items for the test questions. The items in the questionnaire portion were descriptions of tasks in the four subject matters and are called "perceived needs." The items in the test portions covered basic knowledge in measurement and evaluation corresponding to the four subject matters that teachers should know, according to the curriculum of the Measurement and Evaluation workshop program during 1966-1968. These test items are called "actual needs." The four areas are:

- 1) planning test construction or curriculum analysis;
- 2) test item editing;
- 3) item analysis;
- 4) reporting test scores.

Both test and questionnaire were constructed with eight items in each part, with a total of sixty-four items (see Appendix A). The instrument was then translated into Thai. The translation was checked with five Thai students who were enrolled in the College of Education, Michigan State University.

The resulting instrument was tried out on thirty teachers in Khon Kaen, who were then omitted in selecting the sample. The questionnaire and test were distributed to this group in the last week of February, 1974, and all returns were collected in the first week of March, 1974. All responses were coded and data cards were punched. The computer programs were run to compute reliabilities,

item analysis, inter-item correlations, and correlations between actual and perceived needs.

For the perceived needs scale, the reliability of .87 by Hoyt's procedure was obtained. The inter-item correlation was evaluated. One item from each part of the perceived needs scale was deleted. The final form was composed of twenty-eight items, seven for each area of the tasks in measurement and evaluation (see Appendix B.). Both inter-item correlations and variance of items were used as criteria for deletion. Among eight items in Test Planning, items 3 and 4 had negative correlations with some of the other items. Item 3 correlations ranged from $-.236$ to $.451$ and item 4 from $-.246$ to $.444$ (the highest one with item 3). The variance of item 3 was higher than the variance of item 4 (.913 and .907 respectively). Therefore, item 4 was deleted. Among the items in Item Editing, item 16 was taken out. Its correlations ranged from $-.252$ to $.334$ with four out of the seven correlations negative. In the Item Analysis area, item 18 was deleted because of its two negative correlations with item 21 and 22 ($-.007$ and $-.036$ respectively). The lowest inter-item correlations within the last group of items, Reporting Test Scores, were found on item 29 ($.175$ and $.212$). All of the correlations in this group were positive and most of them were higher than .40.

In addition to statistical procedures, the psychometric procedure was also considered. Item 3 and 4 (see Appendix A) in Test Planning were very much alike, i.e., "Deciding what ability should be emphasized in the test" and "Determining the ratio among memory, application, and reasoning." Item 16 of the second group was too general. It was "Writing better items". Therefore, this item was deleted from the

group Item Editing. The next two items were also selected out because they were too general in idea or very broad, i.e., item 18 in Item Analysis group was "Selecting good items", item 29 in Reporting Test Scores group was "Assigning grades to students."

Most of the test items (actual needs) were difficult. The difficulty indices ranged from .06 to .70. Only three items had an index of difficulty above .50. The Hyt's reliability was .41. Three items were deleted from each of the four parts. The deletions were based on the item discrimination index (the point biserial correlation). The deletion was considered separately in each group of measurement areas of the in-service program. Within each of the four subject matters, the item discrimination indices and item difficulties were considered in deleting the items. In the first subject matter, the three lowest discrimination indices were .001, .066, and .153. Though item 3 had a difficulty index of .467 its discrimination index was .001. Item 8 had the highest discrimination index among these three items but had the lowest difficulty index (.100).

In the Item Editing group item 9, 12, and 15 were deleted according to low discrimination indices and low levels of difficulty. Item 9 had discrimination and difficulty indices of -.112 and .267 respectively. Item 12 had .153 and .200. Item 15 had -.137 and .267. The three lowest discrimination indices in the next group, Item Analysis, were .009, .124, and .254 with the levels of difficulty .300, .333, and .167 respectively. In the last group, Reporting Test Scores, the deleted items were items 25, 26, and 29 with discrimination

indices of $-.098$, $-.126$, and $-.080$ and levels of difficulty were $.167$, $.033$, and $.100$ respectively.

Those items retained were then revised. All alternative choices were reordered and reworded if necessary. One alternative among the five-choice items was deleted. The deletion was of the alternative that was the least attractive, i.e., no one chose it or the discrimination index was high and positive. All items then presented four choices, in order to reduce the difficulty of the test items for this particular group of teachers.

The final actual needs test contained twenty items, five for each part (see Appendix B). The correlation between the scores for total perceived needs and for total actual needs was very low ($r = .126$), and not significant.

Data Collection

The revised edition of the questionnaire along with the demographic items and letter (see Appendix B) was sent to teachers in the sample in the last week of May, 1974, both by mail and personal delivery. All teachers in the university area received personal delivery, school by school. Most teachers outside the university area received mail questionnaires. However, some of those who were closest to the university area got theirs by personal delivery. To insure a high percentage of response, the follow-up was done weekly either by postcards or personal contact. By the second week of June, 1974, all returns from the four groups in the university area had been received. The number of returns required from the outside the university area was reached, with the help of some personal contact,

by the last week of June, 1974. As was expected, most of the first forty returns from each group in the outside university area were from people nearby amphurs. The farther the school from the amphurs, the less likely that the responses would be returned.

Since the structure of society and the school system in Thailand are very much the same, teachers who graduated in the same level of teacher certificate must have the same program of study. Because of a very strong centralized educational system, their teaching jobs and routine work are also alike. It is reasonable to assume that teachers who hold the same teacher certificate have the same characteristics. That is the distortion of data is less likely to occur when additional number of teachers is drawn in order to reach the desirable number of responses.

Design

The design of this study is described as a 2x2x2 factorial design with four repeated measures each on actual needs and perceived needs. The four areas of knowledge of educational measurement were: planning test construction or curriculum analysis (PL); test item editing (IE); item analysis (IA); and reporting test scores (RE). Table 3.2 presents the lay-out of the three independent variables: locations of schools; levels of teacher education; and teaching experience. The design is crossed and balanced.

TABLE 3.2

DESIGN OF THE THREE INDEPENDENT VARIABLES:
 LOCATION OF SCHOOLS, LEVELS OF TEACHER EDUCATION,
 AND TEACHING EXPERIENCE

Locations	Levels of Education	Experience	M ₁	M ₂	M ₃	M ₄
University Area (Central Amphur)	Cert.	Old n = 40				
		New n = 40				
	High Cert.	Old n = 40				
		New n = 40				
Far From University (Other Amphurs)	Cert.	Old n = 40				
		New n = 40				
	High Cert.	Old n = 40				
		New n = 40				

Old = teachers who had taught at least eight years

New = teachers who had taught less than eight years

On each scale of needs, the following null hypotheses were tested.

1) There is no difference in needs between teachers located in the area near the university and those farther from the university in Khon Kaen province.

2) There is no difference in needs between teachers who hold the teacher certificate or lower and those who hold higher certificates in Khon Kaen province.

3) There is no difference in the needs of the experienced and of the newer teachers in Khon Kaen province.

4) There is no difference in needs among the four subject matters in measurement covered in the study.

5) There is no location by level of teacher education interaction.

6) There is no location by experience interaction.

7) There is no experience by level of teacher education interaction.

8) There is no location by experience by level of teacher education interaction.

9) There is no subject matter by location interaction.

10) There is no subject matter by level of teacher education interaction.

11) There is no subject matter by experience interaction.

12) There is no subject matter by location by level of teacher education interaction.

13) There is no subject matter by location by experience interaction.

14) There is no subject matter by experience by level of teacher interaction.

15) There is no subject matter by location by experience by level of teacher education interaction.

Analysis

First, the correlation between actual needs and perceived needs will be evaluated to discover whether there was any relationship between them. If there is no significant correlation at all, the perceived needs and actual needs will be analyzed separately using the 2x2x2 factorial design with four repeated measures. If there is a relationship, the multiple repeated measures, both actual needs and perceived needs would have been analyzed in the same lay-out.

Before the analysis of repeated measure design can be performed, the variance-covariance matrices of both actual and perceived needs will be tested for the compound symmetry assumption. Box's chi square test of the homogeneity for variance-covariance matrix will be computed on each of the variance-covariance matrices of actual and perceived needs. If there is a relationship between actual and perceived needs, and the homogeneity of variance-covariance for both matrices are the same either accepted or rejected, the multiple repeated measure analysis will employ the multivariate test procedure. If the outcomes of the Box's chi square tests are not the same, separate analysis will be performed on each of the needs scales. All of the hypotheses will be tested at the significant level of .05 each.

Beside these hypotheses testings, descriptive data were provided to serve as a detailed investigation of some relevant

information such as average number of teachers, average number of students, sex, opinion on national testing, grade taught and subject taught.

Summary

A list of all 5,913 teachers in Khon Kaen province was obtained from the Department of Education, Khon Kaen Education Office. The teachers were classified into eight groups according to two categories each, of school location, level of teacher education, and teaching experience. Forty teachers were randomly selected within each group for the total of 320 plus an additional eighty teachers. A questionnaire containing the demographic questions, perceived need scales, and actual need test items was constructed and tried out in March, 1974. The revised edition of the questionnaire was then sent to those 400 teachers in May, 1974. The required sample size of forty in each group for the total of 320 was reached by the end of June, 1974. All responses were coded and data punch cards were produced. The 2x2x2 factorial design with four repeated measures was used to analyze the data on both actual needs and perceived needs. The decision whether univariate or multivariate analysis will be employed depended upon the results of Box's chi square test of the compound symmetry of the variance-covariance matrices on each need. Graphic presentation will be considered when there is the significance of interaction. Descriptive data along with null hypotheses will also be presented.

CHAPTER IV

ANALYSES AND RESULTS

At the end of the first week of July, 1974, all data were coded. Responses to the perceived needs were coded 4,3,2,1,0 corresponding to the rating from highest need to no need at all on each item. Responses to each item of the actual needs were coded 1 to 5, with 1-4 assigned to responses a-d and 5 assigned where the response was omitted. Coding systems for all demographic data were also created and used to record the information. Next, all data cards were punched, and were ready for use on July 16, 1974. A FORTRAN program was written to obtain the 4 subscale scores on each of the needs scales as well as the total scores for both needs. The accuracy of the program was checked against a sampling calculated by hand. Both yielded the same scores. The program also produced a punched card for each individual respondent. Each card contained information on group identification, demographic data, four subscale scores for perceived needs, total score for perceived needs, four subscale scores for actual needs and total score for actual needs.

All programs were executed by the CDC 6500 at the Computer Center, Michigan State University. The canned programs used were PROFILE to get Box's chi square test for homogeneity of variance-covariance matrix; ACT routine in the CISSR system to get the cross-tabulation or contingency table; FORTAP program to compute Hoyt's reliability and item analysis; and FINN program to perform the correlation, cell means and cell standard deviations, and the analysis

of repeated measure design.

Before any analysis was done, a re-evaluation of the instrument was made. The reliabilities of both the perceived and actual needs scales by Hoyt's procedure were .91 and .42 respectively. The low reliability on the actual needs was probably caused by two factors. First, the test items were rather difficult. Positively skewed distribution was found. Many of the low scores were probably obtained through guessing. Another factor was that the test was quite short (twenty items). The split-half (odd-even split) procedure was employed as another method of estimating reliability. This estimated reliability was .47. If the test was doubled in length (assuming parallel content) the reliability would have been .64.

In this chapter, the results of the data analyses are presented. The descriptive data and some demographic results are discussed first. The correlations between actual and perceived needs are given next. The final section deals with the hypotheses of interest as tested by the repeated measure analyses, graphs of interactions and some follow-up or post hoc tests.

Descriptive Data

The description of the sample is provided to compare the composition of the sampled teacher groups with the population that was discussed in the previous chapter. The average number of teachers and average number of students in the schools in the sample are presented in Table 4.1 along with their standard deviations. The distribution of school size is skewed. Most of the big schools are in the central amphur with thirty to seventy teachers and 1100 to 1500 students. The over-all average number of teachers within a school is twenty-nine with an average

TABLE 4.1

AVERAGES AND STANDARD DEVIATIONS OF THE NUMBER OF TEACHERS
AND NUMBER OF STUDENTS IN SAMPLED SCHOOLS CLASSIFIED
ACCORDING TO EIGHT GROUPS OF TEACHERS

Location	Teacher Education	Experience	Number of Teachers		Number of Students	
			$\bar{X}$	S.D.	$\bar{X}$	S.D.
University	Cert.	New	21.85	11.10	730	254.60
		Old	33.45	14.29	1120	430.98
Area	Higher Cert.	New	42.65	18.65	1255	496.11
		Old	46.05	10.89	1457	217.08
Outside University	Cert.	New	18.95	4.65	672	95.37
		Old	18.35	5.58	570	92.54
Area	Higher Cert.	New	21.00	6.60	660	210.91
		Old	20.75	5.92	674	157.70

TABLE 4.2

INTER-CORRELATIONS AMONG NUMBER OF TEACHERS,
NUMBER OF STUDENTS, PERCEIVED NEEDS, AND ACTUAL NEEDS

	NTE	NST	PN	AN
NTE	1.000			
NST	.520	1.000		
PN	-.069	-.145	1.000	
AN	.160	.152	.004	1.000

NTE = Number of teachers in a school

NST = Number of students in a school

PN = Perceived needs

AN = Actual needs

of 892 students, yielding the ratio of 1:31 teacher to students which is lower than 1:33, the population value.

Table 4.2 presents the inter-correlations among four variables: number of teachers in schools, number of students in schools, perceived needs; and actual needs. The number of teachers as well as number of students in schools serve as the index of size of school the teachers are in. The correlation between the number of teachers and number of students in schools is .520 which is very high. All other correlation coefficients are low, especially the correlation of interest, between perceived needs and actual needs, which is .004. The correlation

between size of school and perceived needs is low and negative, between size of school and actual needs is low and positive.

The responses on the descriptive questions are counted according to category of answer choices. Only the frequency count is of interest. Of all 320 teachers, only fifty-one teachers (16 percent) had attended one of the in-service programs on measurement and evaluation given between 1961 and 1973. Of these fifty-one teachers who had participated in the in-service program, twenty-three are from the central amphur, thirty-six hold the higher certificate, and thirty-three have been teaching for at least eight years.

The majority of teachers in the sample do not like the idea of national testing. Sixty-four percent of the teachers (205) responded no to that question. Of the 160 teachers who have taught more than eight years, seventy-seven percent (123 teachers) disagree with the issue and fifty-one percent (eighty-two teachers) from the less experienced group responded no.

The other breakdowns of the total 320 teachers are by sex, grade taught, and subject taught. There are 141 male and 179 female teachers (44% and 56% respectively). Most of the teachers teach in the elementary grades (grades 1 to 4) and teach all subjects.

One hundred nineteen teachers (37%) teach in grade 1 to 4, ninety teachers (28%) in grades 5-7, one hundred three teachers (32%) in grade 8 to 10, and eight teachers (3%) in grades 11-12. Sixty-three percent of 320 teachers teach a particular subject. Twenty-four teachers teach Thai, thirty-one Sciences, thirty-six Mathematics,

forty-four Social Sciences and Humanities, thirty-seven English, and thirty teach Home Economics, Physical Education, Art and Handicrafts, or Agriculture.

Homogeneity of Variance-Covariance Matrices

As was reported earlier, the correlation between the perceived needs and actual needs was very low. More detailed investigation was done to evaluate the correlation between the subscales of the two need scores. For planning and curriculum analysis, the correlation was $-.070$, Test Item Editing $.065$, Item Analysis $.053$, and Reporting Test Scores $.018$. All correlations were low and not significant at the 95 percent level of confidence (an r of $.098$ is required for significance at $.05$ for $df = 400$). Since there is no correlation between actual and perceived needs the testing of hypotheses of interest on a multiple repeated measures design yields the same results as the combination of the tests on each variable, providing that the over-all Type 1 errors are the same.

The test of the compound symmetry assumption was performed twice, once for the four subscales on the actual needs and again on the perceived needs. Table 4.3 provides the variance-covariance matrix as well as the correlation among the four subscales on actual needs. Again all correlations were low, and only two of them above $.10$ ($r_{PL,IE} = .174$ and $r_{IE,IA} = .175$). The Box's chi square test of homogeneity for the pooled variance-covariance matrix is 11.41 at 8 degrees of freedom which is not significant at the 95 percent level of confidence. Since the hypothesis of the compound symmetry is not rejected, the analysis of variance on repeated measures (Mixed Model

repeated measure) is used to test all hypotheses of interest concerning the actual needs.

Table 4.4 presents the variance-covariance and correlation matrices among the four subscales on perceived needs. All correlations are high, ranging from .43 between Test Planning and Reporting Test Scores, to .591 between Item Analysis and Reporting Test Scores. The variances of Test Planning, Item Editing, and Item Analysis are about the same (22.974, 20.816, and 21.980 respectively), but the variance of Reporting Test Score variable was the highest at 32.414. The Box's chi-square test of homogeneity for the pooled variance-covariance matrix on the perceived needs was 48.59 at eight degrees of freedom which was significant at the .05 level of confidence. (The critical value is 15.51). Therefore, the hypothesis of the compound symmetry was rejected. The multivariate test was employed to test the hypotheses concerning the perceived needs.

Repeated Measure Analysis on Actual Needs

Summary data, cell means and cell standard deviations of the four subscales of the actual needs are presented in Table 4.5. To get a general profile of the four dependent variables, the means and standard deviations for all subjects are also presented. The highest mean was on the Item Editing variable (2.328), and Test Planning (2.175) was the next highest. Third was Item Analysis with a mean of 1.706 and Reporting Test Score was the lowest of all at 1.272 mean. Profiles of all four groups of teachers outside the central amphur were similar to the general profile except for a slight difference for the new higher teacher certificate holders in which

TABLE 4.3

VARIANCE-COVARIANCE AND CORRELATION MATRICES OF FOUR TEST SCORES

	TESTS			
	PL	IE	IA	RE
PL	1.092			
IE	.173	1.268		
IA	.120	.237	1.447	
RE	.070	.117	.099	1.088
PL	1.000			
IE	.147	1.000		
IA	.096	.175	1.000	
RE	.065	.099	.079	1.000

PL = Test Planning

IE = Item Editing

IA = Item Analysis

RE = Reporting Test Scores

Box's chi square test of homogeneity for the pooled variance-covariance matrix = 11.41, df = 8.

(Not significant at .05 level).

TABLE 4.4

VARIANCE-COVARIANCE AND CORRELATION MATRICES
OF FOUR PERCEIVED NEEDS SCORES

	PERCEIVED NEEDS			
	PL	IE	IA	RE
PL	22.974			
IE	12.319	20.816		
IA	11.236	11.571	21.980	
RE	11.726	14.265	15.784	32.414
PL	1.000			
IE	.563	1.000		
IA	.500	.541	1.000	
RE	.430	.549	.591	1.000

PL = Test Planning

IE = Item Editing

IA = Item Analysis

RE = Reporting Test Scores

Box's chi square test of homogeneity for the pooled variance-covariance matrix = 48.59, df = 8.

(Significant at .05 level).

TABLE 4.5
MEANS AND STANDARD DEVIATIONS OF FOUR TEST SCORES
(ACTUAL NEEDS)

Location	Teacher Education	Experience	Tests				Total
			PL	IE	IA	RE	
University	Cert.	New	2.20	1.95	1.63	1.05	6.825
		n = 40	0.99	1.11	1.10	0.99	2.745
		Old	2.15	2.03	1.88	1.40	7.450
		n = 40	1.05	1.03	1.28	0.98	2.230
Area	High Cert.	New	1.85	2.48	1.75	1.85	7.925
		n = 40	1.10	1.22	1.17	1.14	2.422
		Old	2.45	2.43	2.13	0.93	7.925
		n = 40	1.04	1.28	1.24	0.83	2.740
Outside University	Cert.	New	2.15	2.58	1.58	1.40	7.700
		n = 40	1.03	1.11	1.13	1.15	2.345
		Old	2.00	2.18	1.10	0.98	6.250
		n = 40	0.99	1.08	0.90	1.05	2.488
Area	High Cert.	New	2.45	2.43	1.63	1.48	7.975
		n = 40	1.18	0.96	1.37	1.13	2.824
		Old	2.15	2.58	1.98	1.10	7.800
		n = 40	0.98	1.20	1.35	1.03	2.584
All Groups			2.175	2.328	1.706	1.272	
			1.045	1.126	1.203	1.043	

New = Teachers who have been teaching for less than eight years.

Old = Teachers who have been teaching at least eight years.

PL = Test Planning

IE = Item Editing

IA = Item Analysis

RE = Reporting Test Scores

group the means on Test Planning and Item Editing were about the same, i.e., 2.45 and 2.43 respectively. All four profiles for the four groups within the central amphur were quite different from each other and from the general profile. All teacher certificate holders showed higher means on Test Planning (2.20 and 2.15) than means on Item Editing (1.95 and 2.03). Large discrepancies were found among the higher teacher certificate holders within the central amphur. For the new teachers, the mean for Test Planning was much lower than the mean on Item Editing (1.85 versus 2.48) and the reversed order between Item Analysis and Reporting Test Score was observed with the values of 1.75 and 1.85. For more experienced higher teacher certificate holders, means for Test Planning and Item Editing were about the same, with the mean for Item Analysis rather close to them (2.45, 2.43, and 2.13 respectively). These three means were quite different from the mean of 0.93 for the Reporting Test Score variable. The cell standard deviations varied from .83 to 1.37.

The tests of all hypotheses concerning the actual needs are presented in Table 4.6. The univariate repeated measure analysis was used. Fifteen hypotheses were tested. Four of them were on the main effects, the other eleven concerned the interactions.

From Table 4.6 seven of the null hypotheses were rejected at $\alpha = .05$. They were:

- the main effect of teacher education ($F = 8.44$, $p < .003$);
- the interaction between school location and teacher experience ($F = 3.878$, $p < .0499$);
- the subject matter main effect ($F = 67.28$, $p < .0001$);
- the subject matter by location interaction ($F = 3.021$,

TABLE 4.6

UNIVARIATE REPEATED MEASURED ANALYSIS ON TEST SCORES

Sources	DF	MS	F	P
Location (L)	1	0.200	0.123	.727
Teacher Education (T)	1	14.450	8.855	.003
Experience (E)	1	1.250	0.766	.382
LT	1	0.078	0.048	.827
LE	1	6.328	3.878	.0499
TE	1	0.528	0.324	.5691
LTE	1	4.513	2.765	.097
R:LTE	312	1.632		
Subject Matter (M)	3	73.133	67.280	.0001
ML	3	3.284	3.021	<.05
MT	3	1.029	0.946	>.05
ME	3	3.252	2.991	<.05
MLT	3	1.679	1.545	>.05
MLE	3	0.704	0.648	>.05
MTE	3	4.545	4.181	<.01
MLTE	3	4.069	3.743	<.05
RM:LTE	936	1.087		

 $F_{3,1000} (.05) = 2.61$
 $F_{3,1000} (.01) = 3.80$

- $p < .05$);
- the subject matter by experience interaction ($F = 2.991$, $p < .05$);
- the subject matter by teacher education by experience interaction ($F = 4.181$, $p < .01$);
- and, the subject matter by location by teacher education by experience interaction ($F = 3.743$, $p < .05$).

The other nine hypotheses were not rejected, at a 95 percent level of confidence.

The significance of the four-way interaction supported the difference of the profiles among the eight groups of teachers, discussed earlier (under table of means and standard deviations on the actual needs). A follow-up analysis was performed using the same computer program, but the variables were transformed to pairwise comparisons between Test Planning and Reporting Test Scores, Item Editing and Reporting Test Scores, and Item Analysis and Reporting Test Scores. The three-way interaction among location, teacher education, and experience was computed for each transformed variable. The univariate F ratio was significant only for the pairwise comparison of the Test Planning and Reporting Test Scores ($F = 11.0774$ with 1 and 312 degree of freedom, $p < .001$, see Appendix B). The other two comparisons were not significant.

On examining all eight profiles, the most contrasting profiles were from the less experienced higher teacher certificate holders within the central amphur (university area) and the more experienced and higher teacher certificate holders within the central amphur (university area). The first group had the lowest mean on Test Planning

but the highest mean on Reporting Test Scores. On the contrary, the latter group tied for the highest mean on Test Planning but was the lowest on Reporting Test Scores. These reversed profiles suggested that the interaction was not ordinal.

Another follow-up was done to test the three-way interaction among location, teacher education and experience on the Test Planning and the Reporting Test Scores variable as well as the other two variables. The multivariate F ratio was significant at $\alpha = .05$ ($F = 3.618$ with 4 and 309 degrees of freedom, $p < .0068$, see Appendix D) and the univariate F ratio indicated that only the three-way interaction on the Reporting Test Score variable was significant at $\alpha = .0049$. Although there was no significant difference in scores for Test Planning, Item Editing, and Item Analysis among the eight groups of teachers, there were some differences among them on the Reporting Test Score variable, which seems to account for the four-way interaction results.

Since the four-way interaction was significant, a general profile analysis was not made and the following discussion is made on the condition of the presence of these interactions.

The interaction among the subject matter, level of teacher education and experience was significant at $\alpha = .05$ ($F = 4.181$, $p < .01$). It was concluded that the profiles of the four groups of teachers, regardless of the location of their schools, were different, as can be seen in Figure 1. The difference on each of the four subscales on the actual needs between the less experienced and more experienced teachers across teacher certificate holders and higher teacher certificate holders were quite substantial. On

Test Planning, Item Editing and Item Analysis variables, the less experienced groups were higher than those with more experience within the groups of teacher certificate holders. This relationship was reversed within the groups of higher teacher certificate holders. The pattern of differences was not reversed for the Reporting Test Score variable. The differences, rather, were a matter of size. The difference between less experienced and more experienced teachers on the Reporting Test Scores variable was .035 for the teacher certificate holders, and .650 for the higher certificate holders. The differing size of the differences between less experienced and more experienced teachers as well as the direction of differences was also found in the Item Analysis variable, (i.e., .115 within the teacher certificate holders and $-.365$ within the higher teacher certificate holders). Again the analysis of individual variables revealed that the interaction between teacher education and experience was significant only on the Reporting Test Scores variable at $\alpha = .0125$ ($F = 6.895$ with 1 and 312 degrees of freedom, $p < .0091$).

Figure 2 presents the profiles of less experienced teachers and more experienced teachers to explain the interaction between subject matter and experience. The differences between less experienced teachers and the more experienced teachers on the four sub-scales were -0.025 , 0.055 , -0.125 and 0.342 for the Test Planning, Item Editing, Item Analysis and Reporting Test Scores Respectively.

The significance of the interaction between the subject matter and location of school is illustrated by the graph in Figure 3. The differences between teachers in the central amphur and those outside the central amphur on Test Planning, Item Editing,

FIGURE 1

GRAPH PRESENTATION OF CELL MEANS FOR SUBJECT MATTER,
TEACHER EDUCATION, AND EXPERIENCE (ACTUAL NEEDS)

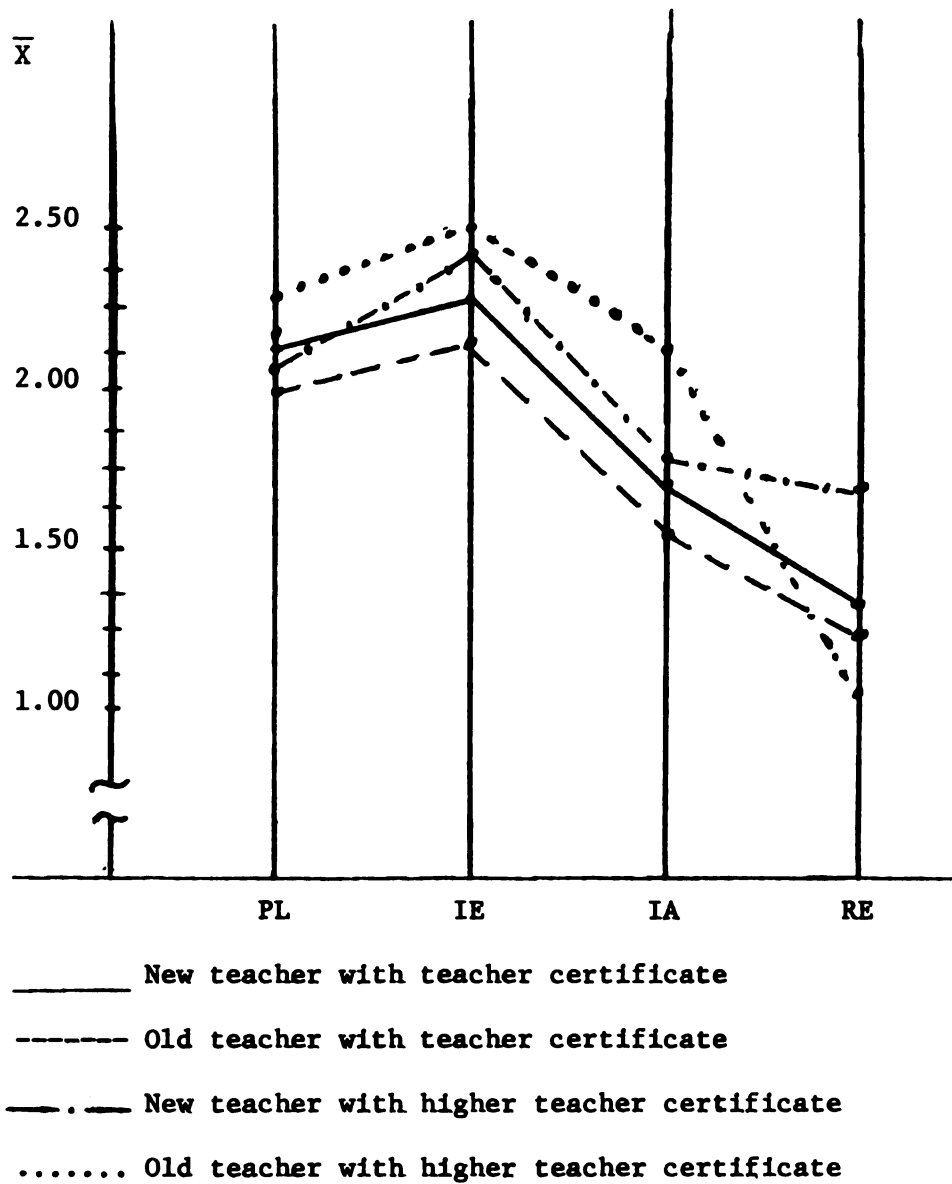
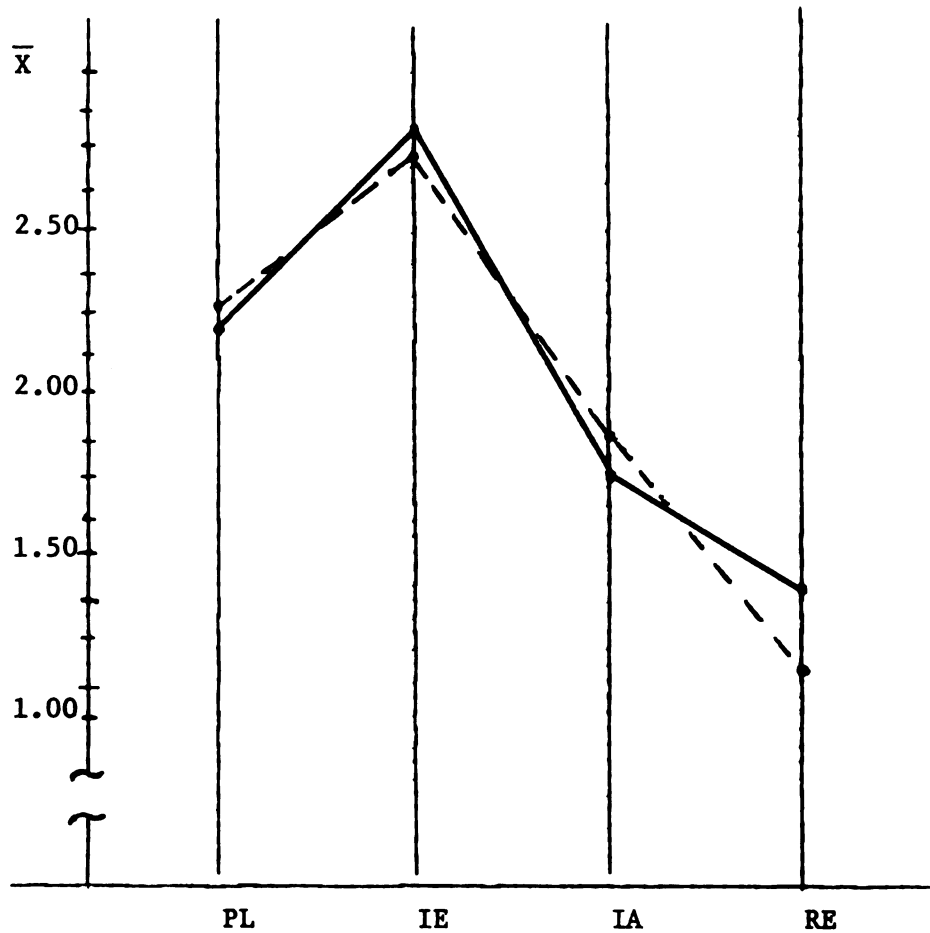


FIGURE 1
(CONTINUED)

		Test Planning (PL)	Item Editing (IE)	Item Analysis (IA)	Reporting Test Scores (RE)
Cert.	New	2.175	2.265	1.605	1.225
	Old	2.075	2.105	1.490	1.190
High	New	2.150	2.455	1.690	1.665
Cert.	Old	2.300	2.505	2.055	1.015

FIGURE 2

GRAPH PRESENTATION OF CELL MEANS FOR SUBJECT MATTER
AND TEACHING EXPERIENCE (ACTUAL NEEDS)



PL = Test Planning

IE = Item Editing New

IA = Item Analysis Old

RE = Reporting Test Scores

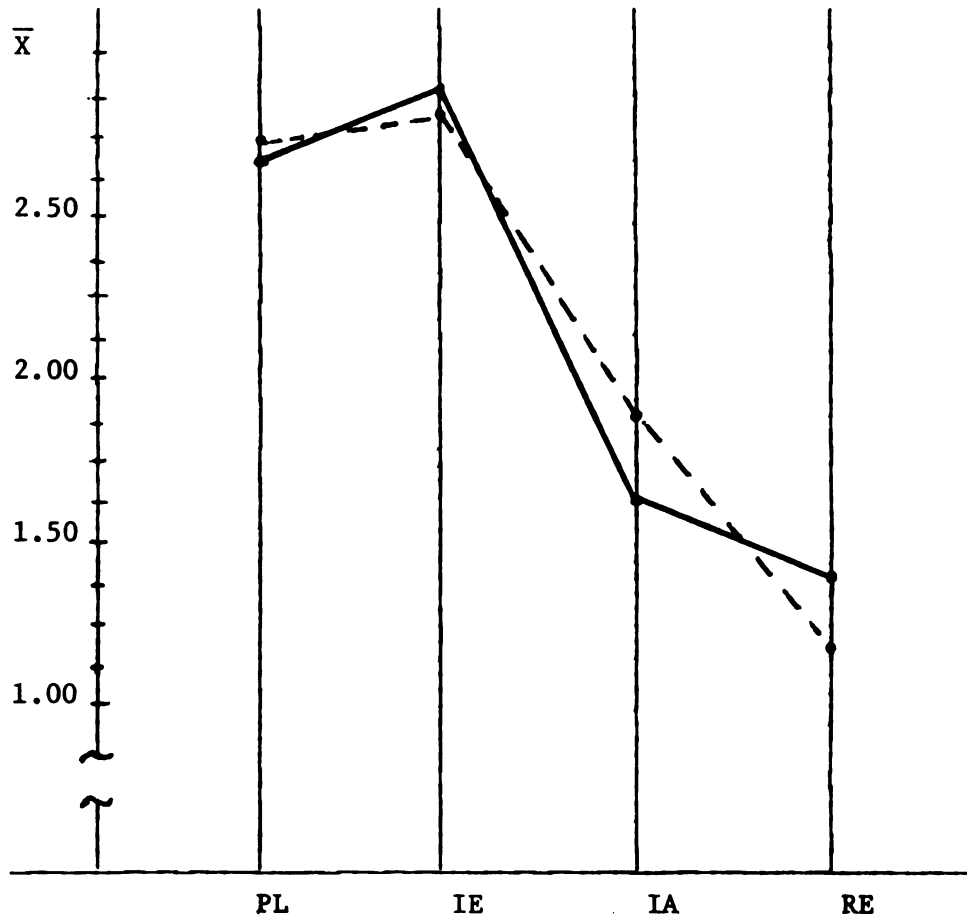
———— New teachers

----- Old teachers

	PL	IE	IA	RE
New	2.163	2.360	1.648	1.445
Old	2.188	2.305	1.773	1.103

FIGURE 3

GRAPH PRESENTATION OF CELL MEANS FOR SUBJECT MATTER
AND LOCATION OF SCHOOLS (ACTUAL NEEDS)



PL = Test Planning

IE = Item Editing

IA = Item Analysis

RE = Reporting Test Scores

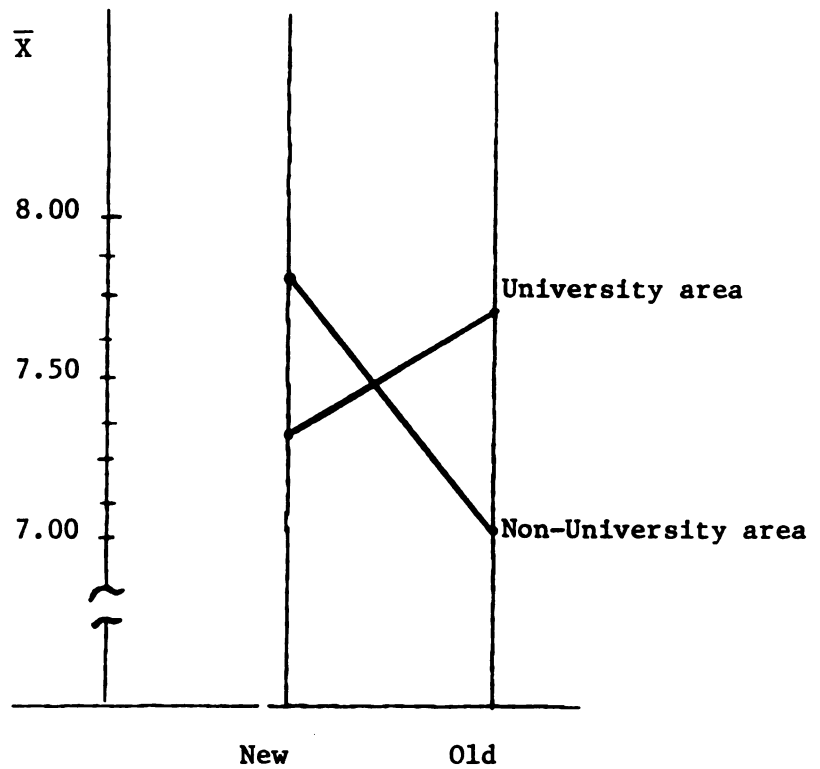
—— University area

- - - - Not in the
University area

	PL	IE	IA	RE
University area	2.163	2.223	1.848	1.308
Non-University area	2.188	2.443	1.573	1.240

FIGURE 4

GRAPH PRESENTATION OF CELL MEANS FOR LOCATIONS OF SCHOOLS
AND TEACHING EXPERIENCE (ACTUAL NEEDS)



New = Teachers who have been teaching for less than 8 years

Old = Teachers who have been teaching 8 years or more

	New	Old
University area	7.375	7.688
Non-University area	7.838	7.025

Item Analysis and Reporting Test Scores were -0.025, -0.220, 0.093 and 0.068 respectively. The interaction was disordinal with respect to location of schools.

On the between-subject part, a significant interaction between location of school and experience was observed. This interaction was explained by the reversed differences in the total actual needs (Fig. 4). The difference between less experienced and more experienced teachers within the central amphur was -0.313 (means of 7.375 and 7.688) while the same difference for teachers outside the central amphur was 0.813 (means of 7.838 and 7.025). The significance of the teacher education main effect was also observed ($F = 8.855$, $p < .003$). The mean for teacher certificate holders was smaller than the mean for higher teacher certificate holders (7.056 and 7.906 respectively).

Repeated Measure Analysis on Perceived Needs

The results of the data analysis on perceived needs are presented in the same pattern as in the previous section. First the summary data are discussed, then the testing hypotheses are presented followed by the graphic presentation of any interactions that were significant.

All summary data, means and standard deviations, both of subscale scores and total perceived needs are presented in Table 4.7. The general profile illustrates all group means on Test Planning (20.816), Item Editing (20.906), Item Analysis (22.606), and Reporting Test Scores (20.897). The maximum perceived needs on Item Analysis with the three other means were about the same. Within the central

TABLE 4.7

MEANS AND STANDARD DEVIATIONS OF FOUR PERCEIVED NEEDS SCORES

Location	Teacher Education	Experience	Needs				Total
			PL	IE	IA	RE	
University	Cert.	New	20.90	20.53	23.05	21.08	85.550
		n = 40	5.07	5.43	4.27	5.00	15.671
		Old	21.93	21.40	23.68	21.73	88.725
		n = 40	4.56	4.06	4.21	6.21	15.320
Area	High Cert.	New	20.23	19.73	22.78	21.10	83.825
		n = 40	5.24	5.43	5.36	6.06	19.339
		Old	21.60	21.20	23.03	21.28	87.100
		n = 40	3.59	4.18	4.22	5.82	14.604
Outside University	Cert.	New	20.65	21.40	21.98	21.65	85.675
		n = 40	5.23	4.73	4.33	4.49	15.100
		Old	19.83	19.40	20.93	18.95	79.100
		n = 40	5.67	4.40	5.98	7.61	18.980
Area	High Cert.	New	21.30	21.68	23.18	21.45	87.600
		n = 40	4.17	3.75	3.78	4.39	12.131
		Old	20.10	21.93	22.25	19.95	84.225
		n = 40	4.48	4.22	4.96	5.28	14.591
All Groups			20.816	20.906	22.606	20.897	
			4.793	4.563	4.688	5.693	

PL = Test Planning

IE = Item Editing

IA = Item Analysis

RE = Reporting Test Scores

New = Teachers who have been

teaching less than eight years

Old = Teachers who have been

teaching eight years or more

amphur (university area), the four groups had almost the same profile. There were small decreases from Test Planning to Item Editing and increases from Item Editing to Item Analysis. The means for Reporting Test Scores dropped to about the same as Test Planning. On the other hand, for those teachers outside the central amphur (outside university area), the means for three groups (except experienced teacher certificate holders) increased from Test Planning to Item Editing and increased to Item Analysis. Finally, for the Reporting Test Scores variable, the means decreased to about the same as the first two subscales. For experienced teacher certificate holders outside the central amphur the profile was the same as for those groups within the central amphur (university area).

Contradictory information between the total actual needs and the total perceived needs was found on the more experienced teacher certificate holders outside the central amphur (outside university area). These teachers had the lowest means of the perceived needs of all eight groups, but they had the lowest actual needs score which indicates that they had the highest need. The group that reported the highest perceived needs was the more experienced teacher certificate holders in the central amphur (mean of 88.725).

The tests of fifteen hypotheses concerning the perceived needs were performed by the multivariate repeated measure analysis. The between-subject part, i.e., hypotheses about location, teacher education, experience, and interactions among them, was tested by the univariate F ratio while the within-subject part, i.e., hypotheses about the subject matter and interactions with it, by the multivariate F ratio.

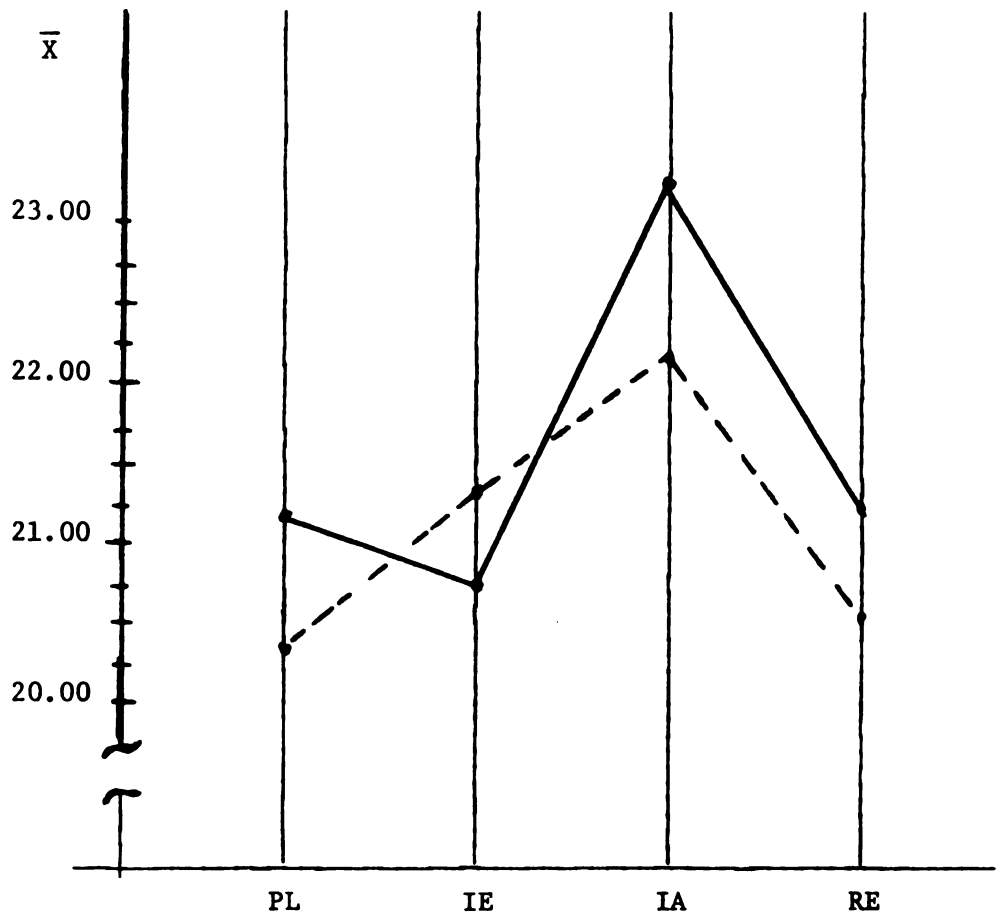
TABLE 4.8

UNIVARIATE AND MULTIVARIATE REPEATED MEASURED ANALYSIS ON PERCEIVED NEEDS

Sources	DF	Univariate			Multivariate	
		MS	F	P	F	P
Location (L)	1	92.450	1.468	.227		
Teacher Education (T)	1	17.112	0.272	.603		
Experience (E)	1	15.313	0.243	.622		
LT	1	135.200	2.146	.144		
LE	1	336.200	5.337	.022		
TE	1	13.613	0.216	.642		
LTE	1	12.013	0.191	.663		
R:LTE	312	62.997				
Subject Matter (M)	3-310				24.349	.0001
ML	3-310				3.351	.0194
MT	3-310				0.458	.7120
ME	3-310				1.115	.3433
MLT	3-310				0.763	.5155
MLE	3-310				0.414	.7434
MTE	3-310				1.082	.3571
MLTE	3-310				0.552	.6474

FIGURE 5

GRAPH PRESENTATION OF CELL MEANS FOR LOCATION OF SCHOOLS
AND SUBJECT MATTER (PERCEIVED NEEDS)



PL = Test Planning

IE = Item Editing

IA = Item Analysis

RE = Reporting Test Scores

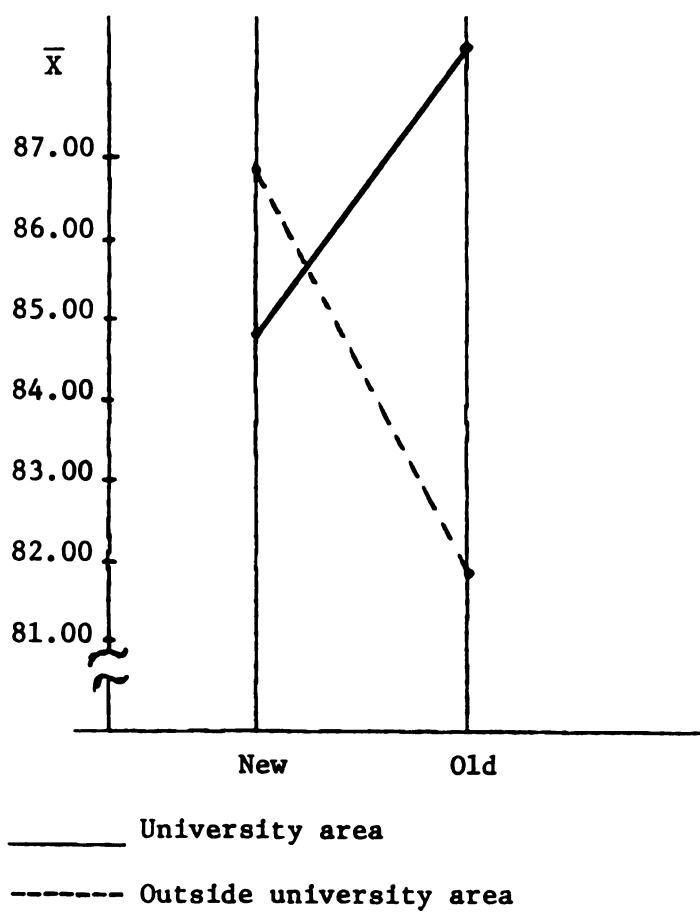
— University area

- - - Outside university

	PL	IE	IA	RE
University area	21.165	20.715	23.135	21.298
Outside university area	20.470	21.103	22.085	20.500

FIGURE 6

GRAPH PRESENTATION OF CELL MEANS FOR LOCATION OF SCHOOLS
AND EXPERIENCE OF TEACHERS (PERCEIVED NEEDS)



	New	Old
University area	84.688	87.913
Outside university area	86.638	81.663

All univariate F and multivariate F ratios are presented in Table 4.8. A rejection of the null hypotheses was indicated, at $\alpha = .05$, on three of the fifteen hypotheses: the interaction between location of schools by experience ($F = 5.377$, with 1 and 312 degrees of freedom, $p < .022$); the subject matter main effect; and the subject matter by location of school interaction (multivariate F of 3.35 and 24.349 respectively with 3 and 310 degrees of freedom, $p < .0001$).

Since the interaction between location and subject matter was significant, the general profile was not interpreted. One profile was made for all teachers in the central amphur (university area) and another for those outside the central amphur (outside university area). Figure 5 presents these two profiles. For teachers in the central amphur (university area), the means vary from 21.165 for Test Planning to 20.715 for Item Editing to 23.135 for Item Analysis and finally to 21.298 for Reporting Test Scores. For teachers outside the university area the mean scores were 20.470, 21.103, 22.083 and 20.50. The largest difference between the groups within and those outside the university area was found on Item Analysis subscale with the value of 1.05.

The interaction between locations of schools and experience was explained by the difference between less experienced teachers and more experienced teachers within the central amphur and the difference for teachers outside the central amphur on the total perceived needs score. The mean for less experienced teachers in the university area was 84.688 and more experienced teachers 87.913 with the difference of -3.225. For teachers outside the university area the mean for the

less experienced group was 86.638 and the more experienced group 81.663 with the difference of 4.975. The type of interaction was disordinal (Fig. 6).

Additional Analysis

Besides testing the hypotheses of interest, further analyses were done to compare actual and perceived needs among various groups of teachers, defined by the following independent variables: having attended the in-service program; sex; favoring a national testing program; and grade taught. The analyses were done individually on both total scores and subscale scores. Means of the groups, F ratios testing the difference between groups, and the probabilities of the F ratios are presented in Tables 4.9 to 4.12.

In Table 4.9, the comparison between teachers who had attended the in-service training program and those who had not attended was made. For total perceived needs, teachers who attended the in-service training had perceived themselves as having a greater need than those who had not attended the training program but the amount of difference was not large enough to be significant at $\alpha = .05$ level. On all four subscales within the perceived needs, no significant differences were found. In the test of actual needs, the teachers who had not attended the in-service training program had higher need (indicated by lower scores from the test) than those who had attended the training program. Again, the amount of the difference was not large enough to show any significance. However, by examining the subscales separately, the difference in Item Editing was significant at $\alpha .05$ in favor of those who had not attended the in-service training

TABLE 4.9

COMPARISON ON PERCEIVED AND ACTUAL NEEDS BETWEEN TEACHERS
WHO ATTENDED AND DID NOT ATTEND THE IN-SERVICE PROGRAM

Variables	In-Service		F P less than	
	Attended n = 51 ($\bar{X}$)	Did Not Attend n = 269 ($\bar{X}$)		
Perceived Needs				
PL	20.71	20.84	0.03	.8588
IE	21.39	20.81	0.68	.4110
IA	22.35	22.65	0.18	.6758
RE	21.94	20.70	2.04	.1541
Total	86.39	85.00	0.32	.5694
Actual Needs				
PL	2.25	2.15	0.35	.5545
IE	2.63	2.27	4.25	.0402
IA	1.78	1.69	0.25	.6203
RE	1.33	1.26	0.20	.6563
Total	8.00	7.38	2.44	.1195

PL = Test Planning

IE = Item Editing

IA = Item Analysis

RE = Reporting Test Scores

TABLE 4.10

COMPARISON ON PERCEIVED AND ACTUAL NEEDS BETWEEN TEACHERS
WHO WERE FAVORABLE AND NOT FAVORABLE TO THE NATIONAL TESTING POLICY

Variables	National Testing		F P less than	
	Favored n = 115 (X)	Not Favored n = 205 (X)		
Perceived Needs				
PL	21.68	20.33	5.91	.0157
IE	21.33	20.67	1.53	.2167
IA	23.00	22.39	1.26	.2630
RE	21.28	20.68	0.80	.3710
Total	87.29	84.07	3.02	.0832
Actual Needs				
PL	2.05	2.24	2.46	.1175
IE	2.43	2.27	1.58	.2095
IA	1.68	1.72	0.09	.7600
RE	1.45	1.07	5.13	.0242
Total	7.62	7.40	0.49	.4828

PL = Test Planning

IE = Item Editing

IA = Item Analysis

RE = Reporting Test Scores

TABLE 4.11

COMPARISON ON PERCEIVED AND ACTUAL NEEDS
BETWEEN MALE AND FEMALE TEACHERS

Variables	Sex		F P less than	
	Male	Female		
	n = 141 ($\bar{X}$)	n = 179 ($\bar{X}$)		
Perceived Needs				
PL	20.30	21.22	2.88	.0908
IE	20.81	20.98	0.11	.7362
IA	22.63	22.59	0.71	.9331
RE	20.90	20.89	0.00	.9916
Total	84.65	85.68	0.33	.5647
Actual Needs				
PL	2.14	2.20	0.25	.6172
IE	2.37	2.30	0.32	.5710
IA	1.84	1.60	3.21	.0742
RE	1.33	1.22	0.83	.3641
Total	7.69	7.32	1.60	.2063

PL = Test Planning

IE = Item Editing

IA = Item Analysis

RE = Reporting Test Scores

TABLE 4.12

COMPARISON ON PERCEIVED AND ACTUAL NEEDS AMONG TEACHERS
WHO TAUGHT DIFFERENT GRADE LEVELS

Variables	Grade Taught				F	P less than
	G ₁	G ₂	G ₃	G ₄		
	n = 119 ($\bar{X}$)	n = 90 ($\bar{X}$)	n = 103 ($\bar{X}$)	n = 8 ($\bar{X}$)		
Perceived Needs						
PL	21.08	21.59	19.91	19.87	2.25	.0830
IE	20.85	21.56	20.61	18.25	1.65	.1784
IA	22.50	23.19	22.17	23.25	0.83	.4789
RE	20.27	22.11	20.58	20.62	1.97	.1185
Total	84.70	88.44	83.27	82.00	1.91	.1284
Actual Needs						
PL	2.11	2.29	2.16	2.12	0.52	.6665
IE	2.13	2.21	2.57	3.37	5.54	.0011
IA	1.47	1.87	1.83	1.87	2.39	.0689
RE	1.22	1.22	1.31	2.12	1.91	.1281
Total	6.93	7.59	7.86	9.50	4.32	.0053

PL = Test Planning

IE = Item Editing

IA = Item Analysis

RE = Reporting Test Scores

G₁ = Taught grades 1-4G₂ = Taught grades 5-7G₃ = Taught grades 8-10G₄ = Taught grade 11 and higher

program. The mean for teachers who had attended the program was 2.63 and for those who had not attended it was 2.24. The F ratio was 4.25 with a probability less than .04. The other three subscales showed no significant differences.

The comparison between teachers who favored a national testing policy and those who did not is presented in Table 4.10. In both the total perceived needs and total actual needs, there were no significant differences between the two groups. However, when the four subscales within each need were analyzed, significant differences were found in the scores for Test Planning in the perceived needs and for Reporting Test Scores in the actual needs. The perceived needs in Test Planning for teachers who favored the national testing policy was higher than for those who did not (means were 21.68 and 20.33 respectively, $F = 5.91$ and $p < .0157$). Conversely, the actual needs in Reporting Test Scores were higher for teachers who did not favor the national testing policy than those teachers who did (means were 1.07 and 1.45 respectively, $F = 5.13$ and $p < .0242$).

A similar comparison was made between male and female teachers. No significant differences were reported in any of the subscales or in the total scores for both perceived and actual needs. A slight difference was found on the total perceived needs with the higher scores made by female teachers. Female teachers had lower scores in the actual needs, but these were not low enough to be significant. Group means on all needs, along with their associated F ratios and probabilities are presented in Table 4.11.

When teachers were compared according to the grades taught, a highly significant difference was observed in the Item Editing actual needs. Those who taught in grades 11 and higher had the lowest actual needs (means of 3.37) and next to the lowest were teachers who taught in grades 8 to 10. The actual needs in Item Editing increased as the level of grade taught decreased. Teachers who taught in grades 1 to 4 had maximum needs (mean 2.13). The F ratio was 5.54 which was significant at $\alpha = .01$ (see Table 4.12). This finding agrees with the previous hypothesis which indicated that teachers who had higher teacher education had less actual need. Since most of the teacher certificate holders or lower were teaching in the lower levels, it is to be expected that teachers in grades 1 to 4 should show the highest actual needs. The pattern of differences among the four groups of teachers, classified by the grade level they taught, was also reflected in the scores for the total actual needs. Means were 6.93, 7.59, 7.86, and 9.50 for those who taught in grades 1 to 4, 5 to 7, 8 to 10, and 11 and higher, respectively. The F ratio was 4.32 with probability less than .0053.

Summary

The correlation between the actual and perceived needs was very low and homogeneity of the variance-covariance matrix for the actual needs was accepted but rejected for the perceived needs. The decision was made that the analyses should be separated between actual and perceived needs. The first descriptive discussion was on information from the demographic items. The data obtained tended to support that the sample represented the population. The univariate

repeated measure analysis was employed to test hypotheses concerning the actual needs. The four-way interaction was significant. The general profile was not interpreted. Summary data indicated that the more experienced teacher certificate holders outside the university area had the highest actual needs scores. Among the subject matter Reporting Test Scores was the most needed area. Item Analysis was next. The teacher education main effect was also significant. The means indicated that the teacher certificate holders needed help more than the higher teacher certificate holders.

The multivariate repeated measures analysis was employed to test hypotheses concerning the perceived needs. The subject matter by location of school interaction was significant. Interpretations of profiles were made separately for the central amphur teachers and for those outside. Both groups had the most need in Item Analysis. Teachers in the central amphur were higher than those outside the university area on their perceived needs on the Item Analysis variable. In the university area, Item Editing was the least needed but in the outside group Test Planning was the least needed. Disordinal type interaction was found on the location of school by experience interaction. Means of the total perceived needs indicated that the more experienced teacher certificate holders outside the university area had the least perceived needs, which was in contradiction with the findings on actual needs.

The hypotheses testing results were as follow:

- Actual needs

- 1) There was no difference in actual needs between teachers in and outside of the university area.

2) Teacher certificate holders had more actual needs than the higher teacher certificate holders.

3) There was no difference in actual needs between more and less experienced teachers.

4) There were some difference in actual needs among the four subject matters; Reporting Test Scores seemed to be the most needed (lowest mean).

5) There was no interaction between location of schools and level of teacher education.

6) There was a disordinal interaction between location of schools and teaching experience. Less experienced teachers in the university area had more actual needs than less experienced teachers outside the university area but the more experienced teachers in the university area had less actual needs than the more experienced teachers outside the university area (Fig. 4).

7) There was no interaction between teaching experience and level of teacher education.

8) There was no three-way interaction among location of schools, teaching experience, and level of teacher education.

9) There was a disordinal interaction between subject matter and location of schools. Test Planning and Item Editing variables both in and outside the university areas had about the same needs. For Item Analysis, the university area group had more need than the outside university area group but the effect was reversed in the Reporting Test Score variable.

10) There was no interaction between subject matter and level of teacher education.

11) There was a disordinal interaction between subject matter and teaching experience. The less experienced teachers seemed to have less need in Reporting Test Score.

12) There was no interaction among subject matter, location of schools and level of teacher education.

13) There was no interaction among the subject matter, location of schools and teaching experience.

14) There was a three-way interaction among the subject matter, teaching experience, and levels of teacher education (Fig. 1).

15) There was a four-way interaction among the subject matter, location of schools, teaching experience, and level of teacher certificate. The follow-up test observed that there was no difference in scores for Test Planning, Item Editing, and Item Analysis among eight groups of teachers, but there were some differences among them on the Reporting Test Scores variable.

- Perceived needs

1) There was no difference in perceived needs between teachers in and outside the university area.

2) There was no difference in perceived needs between the two groups of teachers classified by teacher education levels.

3) There was no difference in perceived needs between more and less experienced teachers.

4) There were some differences in perceived needs in the four subject matters. The Item Analysis variable seemed to be the most needed while the other three were about the same.

5) There was no interaction between location of schools and level of teacher education.

6) There was a disordinal interaction between location of schools and teaching experience. Among the less experienced teachers, those who were in the university area had fewer needs than the outside university area, but for more experienced teachers, outside university area had fewer needs than the university area.

7) There was no interaction between teaching experience and level of teacher education.

8) There was no interaction among location of schools, teaching experience, and levels of teacher education.

9) There was an interaction between the subject matter and location of schools as presented in Figure 5. The university area group had more needs in the Test Planning, Item Analysis, and Reporting Test Score variables but had less need in the Item Editing variable.

10) There was no interaction between the subject matter and level of teacher education.

11) There was no interaction between the subject matter and teaching experience.

12) There was no interaction among the subject matter, location of schools, and levels of teacher education.

13) There was no interaction among the subject matter, location of schools, and teaching experience.

14) There was no interaction among the subject matter, teaching experience, and level of teacher education.

15) There was no four-way interaction among the subject matter, location of schools, teaching experience, and levels of teacher education.

Analyses were also done to compare actual and perceived needs among various groups of teachers by the following independent variables: having attended the training program; sex; favoring a national testing program; and grade taught.

Among teachers who attended and did not attend the in-service program, the former group had perceived themselves as having a greater need but had lower actual needs (indicated by higher scores from the test) than those who had not attended the training program. There were no significant differences on needs among teachers who favored and did not favor the national testing policy. Neither were there any differences between the sexes.

There were differences among teachers who taught a different grade. Teachers who taught in higher grades had lower needs than those who taught in lower grade levels.

CHAPTER V

SUMMARY AND CONCLUSIONS

In-service programs in educational measurement, with emphasis on test construction procedures, i.e., test planning and curriculum analysis, test item editing, item analysis, and reporting test scores, have been given to teachers in various parts of Thailand by the Ministry of Education and the Test Bureau at the College of Education, Prasarnmitr. Undoubtedly, the cost of this in-service program was high and the teachers who joined the program might not be those who actually needed the most help. It was the purpose of this study to find out which groups of teachers actually need in-service training and in which areas of measurement the need is the greatest. This study also yields some follow-up information on the effects of the earlier in-service programs. Two types of needs, actual and perceived, were investigated.

Summary and Discussion

The data of this study were based on two factors: a) the perceived needs which were measures of the value judgment of teachers in the sample; and b) the actual needs which were measures of their knowledge in measurement and evaluation. A Likert type questionnaire was constructed to measure the perceived needs. The items were a random sample of the tasks usually involved in the process of test construction. The respondent was asked to rate each task from 1 to 5 where 1 meant most needed and 5 not needed at all. The first edition of the questionnaires had eight items for each of the areas of in-service training for a total

of thirty-two items. The actual needs were measured by a random sampling of test items measuring knowledge on educational measurement similar to those items used in basic measurement courses at Michigan State University. There were thirty-two items, eight each under the four areas of the in-service program. Both questionnaire and test were tried out and the reliabilities were found to be .87 and .41 respectively.

The revised editions of the questionnaire and test had twenty-eight items and twenty items respectively. They were sent together with a demographic data sheet and introduction letter to 400 teachers who were randomly selected from eight strata in Khon Kaen province. The stratification was based on the three variables: location of school-university area or out of university area; levels of teacher education-teacher certificate holders or higher teacher certificate holders; and teacher experience-less experience or more experience using eight years as the dividing line. Responses were received from teachers forming an equal sample size for each of the eight strata.

The correlation between total actual needs and total perceived needs was very low, i.e., .004. It was indicated that there was no relationship at all between the actual and the perceived needs. Since most of the teachers had little knowledge about the subject matter, they felt a great need in every task. At a particular level of actual needs, there were some teachers who would like to know more about the subject matter and some who reported lower perceived needs. For the latter group, the perceived needs were compared with routine work in school and teachers seemed to decide that they had no time to pay attention to measurement problems. At the level of their present knowledge on the subject matter, they seemed to think that they could carry on test construction and measurement in general very well. The variation in their

perception of their needs among those teachers who had the same level of actual needs caused the low correlation between the two needs. Another way to explain the low correlation between perceived and actual needs is by the observation that marginal distributions of both variables were skewed. Most of the teachers reported high perceived needs and low actual needs. The scatter diagram formed a tendency of curvilinear relationship between the two variables. Therefore, the Pearson product moment correlation yielded a low value.

From this study, the perceived needs were the teachers' value judgments about their needs in measurement and evaluation while the actual needs were what they knew in the subject matter areas. Since value judgments were different among those teachers who had the same levels of actual needs, this also contributed to the low correlation coefficient between the two needs.

The correlation between the number of teachers in the respective respondents' school with their perceived needs was low and negative at $-.069$ and with their actual needs was low at $.160$. Since there was a high relationship between the number of teachers in school and the size of schools, the low correlation observed indicated that there was no relationship between the size of schools and the actual needs and between the size of schools and the perceived needs. The size of school was also related to the number of grades in school. Big schools generally were those that had grades 1 to 12. No matter how big the schools the teachers were in, their jobs were still determined by their levels of teacher education. Big schools included all levels of teachers, from those who had no teacher education at all to those who had master's degrees. The variation of teacher certificate in the big schools was greater than the variation in small schools.

The test of homogeneity of the variance-covariance matrix was not significant for the four subscales of the actual needs but was significant for

the perceived needs. The univariate repeated measures analysis was applied to the actual needs and the multivariate test to the perceived needs. The three way factorial design (2x2x2) with four repeated measures were analyzed by the FINN program. The design was crossed and balanced.

The analysis of the actual needs showed that the interaction among location of schools, level of teacher education, teaching experience, and the subject matter was significant. Descriptive follow-ups showed that the less experienced teachers holding only the teacher certificate who worked within the university area and the more experienced holders of the higher teacher certificate who also worked within the university area had the most contrasting profiles. The four way interaction was explained by the significance of the location by teacher education by teacher experience interaction on the contrast variable of Test Planning and Reporting Test Scores. The three-way interaction was not significant when Test Planning was analyzed individually, but the interaction was significant for Reporting Test Scores alone. In the presence of the four-way interaction, the subject matter by the level of teacher education by experience interaction, the subject matter by experience interaction, the subject matter by location interaction, the subject matter main effect, the location by experience interaction, and the teacher education main effects were significant at $\alpha = .05$. All interactions were presented in graph form.

The F ratio for testing two hypotheses concerning the main effects were significant. They were the subject matter main effect and the teacher education main effect. On the general profile actual needs in Reporting Test Scores was the highest need (lowest scores from the test) followed by Item Analysis and by Test Planning with Item Editing the lowest. However the general profile could not be made applicable to

all groups of teachers because the four-way interaction was significant. The significance of the teacher education main effect was followed by the notion that the teacher certificate holders had more actual needs than the higher teacher certificate holders.

On perceived needs, only three null hypotheses were rejected. The subject matter by location interaction was studied along with the subject matter main effect. For teachers in the university area, the more experienced teachers had perceived needs ranging from highest to lowest in the following order: Item Analysis; Reporting Test Scores; Test Planning; and Item Editing. But for those outside the university area, the perceived needs on Item Editing was second, followed by Reporting Test Scores and Test Planning. In general, the factor of Item Analysis had the highest scores in perceived needs. The other null hypothesis rejected was the location of schools by teaching experience interaction. The interaction was disordinal which could be explained by the observation that less experienced teachers within the university area had lower perceived needs than the more experienced teachers within the university area but less experienced teachers outside the university area had higher perceived needs than more experienced teachers outside university area. The difference between perceived needs of all (both inside and outside the university area) of the less experienced teachers was smaller than all of the more experienced teachers.

Further comparisons were made to find whether there was any difference among levels of the following independent variables: having attended the in-service program; sex; favoring a national testing program; and grade taught. On total perceived and actual needs, a difference was found on the grade taught variable for total actual needs only.

It was indicated that the higher the grade taught, the higher the test scores and hence the lower their actual needs. The same observation was true for actual needs in Item Editing but not for the other three subject matters. The score in actual needs for Item Editing also differed between teachers who had attended the in-service program and those who had not, even though no difference was found in the total needs. The means indicated that those who had attended had lower actual needs.

Between those who favored a national testing program and those who did not, the differences were found on Test Planning perceived needs and Reporting Test Scores actual needs. The means showed that teachers who favored a national testing program had higher Test Planning perceived needs than those who did not. For Reporting Test Scores actual needs, teachers who favored national testing needed less. There were no significant difference at all between male and female teachers.

Conclusions and Implications

The descriptive information supported that the random sample was very similar to the total population. This suggests that the bias due to the substitution of the samples in the outside-university area is minimum. The smaller teacher-student ratio for the sample as a whole was contributed by the larger proportion of central amphur schools (university area) than in the total population. Most of the large schools, all of which were in the university area, had lower teacher-student ratios. Most of the schools outside the university area had much higher ratios because of the difficulty of getting teachers in the remote areas. Also, some of these remote schools are one-room schools.

Sixteen percent of the 320 teachers in the sample had experience

in the in-service training program between 1961 and 1973. Seventy-one percent of these fifty-one teachers held at least the higher teacher certificate, and sixty-five percent were more experienced teachers, who had been teaching for at least eight years. The conclusion could be drawn that in those years the program of in-service training was given to those who had already had more teacher education. Because of the nature of the subject matter, with some mathematics involved, the holders of only the teacher certificate and those teachers who had not had any teacher education at all may have turned down the invitation to join the program.

A cross-tabulation between responses on attended/did not attend the in-service program and favored/did not favor a national testing policy revealed that of fifty-one teachers who did attend the in-service program, thirteen teachers favored a national testing policy. Thirty-eight teachers who attended the in-service program were unfavorable to a national testing policy. Twenty-six percent (thirteen out of fifty-one) of the teachers who attended the in-service program favored a national testing policy. On the other hand, thirty-eight percent (102 out of 269) of the teachers who did not attend the in-service program reported favoring such a policy.

In the past, the in-service training program emphasized item editing in an attempt to improve the quality of the teacher-made test. The results of this attempt were reflected in the lower needs on both the actual and perceived needs tests than were found in other subject matter, i.e., lower than Item Analysis within perceived needs and lower than both Reporting Test Scores and Item Analysis in

actual needs. Because of the interaction of school location, teacher education, teacher experience and subject matters, the needs for each group were different. Therefore, the subject matter should be arranged according to the needs of a majority of teachers in each training session. The results of this study seem to indicate that it would be appropriate to emphasize item analysis procedures in future in-service programs because it was the area with the highest perceived needs score and was next to the highest in the test of actual needs. Further, item analysis is the next step in improving tests after the items have been written and tried out.

In the Item Editing subject matter, teachers in the higher grades had less actual need than those who taught in lower grades. This may suggest that in the past, both in the in-service training program and in teacher-training curriculum, the emphasis was on item editing, i.e., writing the test questions more than any other area of subject matter in measurement and evaluation.

In testing the difference between teachers who had attended and had not attended the in-service program, no significant difference was found between these two groups. But the data showed that there were some slight differences between them. In perceived needs, teachers who had attended the in-service training program felt more need than those who had not attended the program. This result might suggest that those who attended the program seemed to realize that there were several things that they must know about tests and measurement. The data, also, indicated that teachers who had attended the training program had higher scores on the actual needs test

(meaning that they had lower needs) than who had not attended the program. And, again, the significant difference was found in the Item Editing subject matter.

One of the nation-wide programs conducted by Chaval Paeratakul during 1965 included spending two days out of the five-days sessions on test item editing alone. The activities included theory in writing good test items with examples of good and less good items. Then each teacher had to write some test questions and criticize them item by item. Another piece of evidence to support the position that there is less need in Item Editing subject matter was that in Chaval Paeratakul's measurement textbook, the first book in Thai language about measurement, the discussion about item editing subject matter takes four out of the nine chapters.

The study also suggests that there should be a compulsory program requiring the holders of teacher certificates and teachers with no formal teacher education to attend the training sessions, especially those who have been teaching for at least eight years outside the university area and for new teachers within the university area. Special attempts should be made to give service to these two groups of teachers. The high actual needs for these two groups may be the result of a lack of interest and or of heavy loads or of routine teaching. Most of the teachers studied taught in the lower elementary grades (1 to 4), taught all subjects and for about 5 hours a day. They apparently found no time to study or to pay much attention to other professional activities.

Recommendations for Further Study

Since the previous in-service training seemed to yield little benefit to the teacher certificate holders, perhaps because the work may have been too hard and conceptually the subject matter too difficult for them, there should be a study of any future in-service training to find out which group of teachers benefits the most from the training. Another attempt should be made to find suitable ways of presentation and interpretations of subject matter for those who have had no teacher education at all. An experimental study should be done on the effect of the in-service training on actual measurement in the schools. Each year the Department of Education in Khon Kaen has about 120 positions open for graduates from the teacher training institutions. These new teachers might be randomly assigned to training and no-training groups. The no-training group would then be sent directly to the assigned schools. The training group would first go through a training program and then be sent to assigned schools. The follow-up study could then compare the quality of the tests these new teachers made. The efficiency of the training program might be evaluated from the proposed study.

Another experiment might be done to investigate the benefit from the training service. The gain of knowledge after the training should be studied to compare the gain made by holders of the teacher certificate with the gain made by the holders of the higher teacher certificate. The quality of the tests they use should also be investigated to discover the extent to which the new knowledge they have acquired is put into practice, and under what circumstances the application is facilitated.

If there is no improvement in the quality of the tests the in-service training is in vain. The effort and money that was put on the program were wasted.

The application of the test results has been highly important in the Thai educational system. One test result might determine the students' whole future. If the test was not valid and reliable, a capable student might be forced to drop out of school. In order to evaluate students' abilities based on valid and reliable measures, sound national testing programs should be developed and put into practice. Standardized tests should be constructed and be available to all teachers. Funds should be provided by the Ministry of Education to construct national norms. A National Test Bureau should be established, to be a center of testing services and to carry on all educational testing jobs. Programs of in-service training could be part of the function of this National Test Bureau. Different programs should be established for different groups of teachers, depending on their needs and the levels of their education. For example, teachers with minimum educational background should be provided with basic information about how to construct a test, to write good items, and to interpret the test scores, since these subject matters do not require much knowledge of statistical procedures. Item analysis and transforming scores should be tasks done for them by personnel at the local educational offices.

Since the four-way interaction (subject matter by location of schools by level of teacher education by teaching experience) was found, yet needs of teachers in each group were different in some way, different programs of in-service training must be planned for

particular groups of teachers and these programs must be flexible. That is, it is desirable that before planning an in-service program, a pilot study be conducted to find out the needs of teachers in the particular location and the level of their knowledge about the subject.

Besides the standardization of tests and administration of the in-service program, the National Test Bureau should provide an information service, either through the preparation and distribution of a teachers' handbook on test construction, or through a quarterly journal on educational measurement and evaluation, or desirably through both.

The College of Education at Khon Kaen University should add another dimension to respond to teachers' testing needs. As the closest resource, the university can provide an intermediate help to all teachers in the province. The services can be as individualized as mail corresponding, providing a practical suggestion to individual teacher's problems. The university test and research center should offer services such as consultation and information services. Graduate students majoring in Measurement and Evaluation should acquire most valuable experience through participation in this program.

A summer work-shop is another service which the university can provide to teachers. Of course content and structure should be drawn carefully based on participants' characteristics and needs. The session should be about one to one and a half months long. Practical considerations would be more important than a theoretical orientation. Before the participant teachers leave the campus, they should be convinced that they can get help from the staff any time on any measurement problem that they may have in the future. They should leave with the feeling that

the university is trying to help all teachers, both participants and non-participants. The university is the place they can turn to when they would like to have their problems solved.

APPENDIX A

APPENDIX A

THE INSTRUMENT (FIRST EDITION)

Directions: You need not identify yourself, but please answer each of the questions as accurately and as honestly as you can.

PART I. Please mark X in the appropriate columns for each item.

Do you think you need to know more about these tasks?

	Urgent Need	Very Much Needed	Some Need	Little Need	Don't Need
1. Choosing the type of test to use					
2. Determining the length of the test					
3. Deciding what ability should be emphasized in the test					
4. Determining the ratio among questions that call for memory, application, and reasoning.					
5. Determining the number of items for each topic					
6. Administering the test					
7. Setting the difficulty level					
8. Arranging the items into the test					
9. Writing good item stems (in multiple-choice item)					
10. Writing reasonable alternatives (in multiple-choice item)					

11. Writing good true-false items
12. Scoring essay tests
13. How to measure in the cognitive domain
14. How to measure in the affective domain
15. How to measure in the psycho-motor domain
16. Writing better items
17. Understanding the major uses and limitations of item analysis data
18. Selecting good items.
19. Computing and understanding the item difficulty index
20. Computing and understanding the item discrimination index
21. Understanding the difference and relationship between difficulty and discrimination
22. Finding the validity of a test
23. Finding the reliability of a test
24. Revising items
25. Reporting and writing recommendations to the principal and superintendent
26. Choosing the type of test scores
27. Computing and understanding T-scores
28. Computing and understanding percentiles
29. Assigning grades to students

Urgent Need

Very Much Needed

Some Need

Little Need

Don't Need

- 30. Drawing and interpreting profiles
- 31. Using norms
- 32. Distinguishing between norms and standard

	Urgent Need	Very Much Needed	Some Need	Little Need	Don't Need

PART II Mark X in front of the best answer of each item.

1. In evaluating an achievement test for possible use in your school, the most important issue is whether the test
 - a) is based upon a thorough survey of teaching practices over the whole country.
 - b) matches in content the objectives of instruction in the school.
 - c) has sufficiently high predictive validity.
 - d) is convenient to administer and score.
 - e) has sufficiently high reliability.
2. The most important concern in judging how well a test samples a content area is its
 - a) completeness.
 - b) correlation with a criterion.
 - c) representativeness.
 - d) variety.
3. What is the first decision made by the test constructor?
 - a) Characteristics of students taking the test.
 - b) Number of test items to use.
 - c) Types of test items to use.
 - d) Difficulty of the test.
 - e) Purpose of the test.
4. In constructing a table of specifications for a classroom test, the number of test items allotted to each cell should be
 - a) varied each term, so that students are unable to detect the amount of emphasis to be given to each area in the test.
 - b) approximately equal, so that no particular objective will receive undue emphasis.
 - c) balanced between knowledge and higher level learning outcomes.
 - d) determined by the emphasis to be given during instruction.

5. Which of the following will not be accomplished by constructing a test blueprint?
- a) help to clarify instructional objectives.
 - b) improve the validity of the test.
 - c) relate evaluation to teaching.
 - d) help to "balance" the test.
 - e) save construction time.
6. Which one of the test attributes is most likely to be use of a table of specifications?
- a) Difficulty
 - b) Objectivity
 - c) Reliability
 - d) Sampling
 - e) Usability
7. Which type of item format provides greatest content sampling in the shortest possible testing time?
- a) essay
 - b) matching
 - c) multiple-choice
 - d) short-answer
 - e) true-false
8. Which of the following is not a purpose of the classroom achievement test?
- a) Measuring attainment of instructional objectives.
 - b) Diagnosing individual learning difficulties.
 - c) Judging relative mastery of basic knowledge.
 - d) Assessing relative curricular effectiveness.
 - e) Assessing instructional effectiveness.

9. The best source of possible alternatives for a multiple-choice item is the
 - a) explanations of poor students in the class.
 - b) teacher's knowledge of the subject matter.
 - c) questions of bright students in the class.
 - d) textual material assigned to the class.
10. Multiple-choice items written in the form of questions
 - a) assume complex reading skills.
 - b) require no more detailed alternatives than do incomplete statements.
 - c) have greater clarity than items phrased in incomplete statements.
 - d) lend themselves to easier detection of construction flaws.
11. It is good practice in multiple-choice, single-correct answer test items to
 - a) randomize the order of the answers, putting the correct alternative in each position an approximately equal number of times.
 - b) put the correct answer last or next to last more often than in any other position.
 - c) let chance take care of the order of alternatives within the item.
 - d) arrange placement of the correct answers according to some predetermined scheme.
12. "None of the above" is an appropriate distractor for a multiple-choice item in cases where
 - a) the other distractors are partially but not absolutely correct.
 - b) the best answer is an absolutely correct answer.
 - c) the number of possible responses is limited to two or three.
 - d) a large variety of possible responses might be given.
 - e) guessing is apt to be serious problem.

13. Which, if any, of the following should characterize the stem of a good multiple-choice test item?
- a) It should be a complete direct question.
 - b) It should be the subject portion of the proposition the item is intended to test.
 - c) It should express or imply a specific question.
 - d) It need not have any of the foregoing characteristics.
14. How should the question and the responses of a multiple-choice test item be changed to make the item easier?
- a) question more specific, responses more alike.
 - b) question more specific, responses more diverse.
 - c) question more general, responses more alike.
 - d) question more general, responses more diverse.
15. With respect to the multiple-choice test item a good distracter is
- a) only an untrue statement.
 - b) only irrelevant statement.
 - c) both untrue and irrelevant.
 - d) either untrue or irrelevant.
16. Within a given item format, the items of a test should be arranged by
- a) subject-matter content.
 - b) increasing level of complexity.
 - c) level of reading difficulty.
 - d) a random pattern.
17. If experts agree on the correct answer to a test item, the item is
- a) fair.
 - b) objective.
 - c) reliable.
 - d) specific.

18. Fundamentally, item discrimination indices attempt to indicate

- a) the ability of items to distinguish between persons having different levels of that characteristics which the test measures.
- b) how well the items can be distinguished from one another within the test.
- c) the degree to which items from different tests can be discriminated.
- d) the ability of the item writer to develop items which are readable.

19. If three good students out of four answer a question correctly, what is its index of discrimination?

- a) 1.00
- b) .75
- c) .25
- d) Insufficient information is given to answer the question.

20. It is likely that item discrimination will be at a maximum when the item difficulty index approaches

- a) 1.00
- b) .75
- c) .50
- d) .25
- e) 0.00

Items 21 through 24 are based on the following data:

A class of 200 students was divided at the median into a high and a low group. Their responses to five 4 choice multiple-choice items are tabulated below. Correct responses are underlined.

<u>Item Number</u>	<u>Group</u>	<u>Number of students marking</u>			
		<u>Response 1</u>	<u>Response 2</u>	<u>Response 3</u>	<u>Response 4</u>
1	High	<u>80</u>	20	0	0
	Low	<u>45</u>	30	20	5
2	High	0	0	<u>100</u>	0
	Low	0	15	<u>85</u>	0
3	High	<u>35</u>	10	10	45
	Low	<u>55</u>	20	0	25
4	High	20	5	<u>65</u>	10
	Low	20	25	<u>35</u>	20
5	High	30	30	0	<u>40</u>
	Low	5	55	5	<u>35</u>

Use the following KEY to answer each of the next four items:

- KEY:
- a) Item 1
 - b) Item 2
 - c) Item 3
 - d) Item 4
 - e) Item 5

- 21. Which item discriminated best between high and low groups? _____
- 22. Which item has the highest index of difficulty? _____
- 23. Which item has difficulty index of .50? _____
- 24. Which item shows negative discrimination? _____

24. The most easily defensible use of tests in schools is for
- a) accreditation of schools.
 - b) motivation of study.
 - c) teacher evaluation.
 - d) pupil guidance.
26. If a pupil had a score at the median of his class, we know he
- a) is at a score half-way between the lowest and highest score.
 - b) has a score higher than those of half his classmates.
 - c) is doing acceptable work.
 - d) has a score at the mean.
27. Norms are least likely to remain current for which type of test?
- a) Achievement
 - b) Aptitude
 - c) Interest
 - d) Projective
28. On test XYZ pupil A scored at the 80th percentile and pupil B scored at the 40th percentile. What statement may we make relative to their abilities?
- a) Pupil A has twice the measured ability of Pupil B.
 - b) Pupil A's test score will be twice that of Pupil B.
 - c) If Pupil A's score is superior to those of 70 pupils, B's score is superior to those of 35.
 - d) Percentile ranks are meaningless in comparing the scores of A and B.
 - e) We need to know more about the distribution of scores before we can compare the two percentiles.

29. A distribution of linear z -scores is always
- a) similar to the raw score distribution.
 - b) a rectangular distribution.
 - c) a normal distribution.
 - d) a peaked distribution.
 - e) a skewed distribution.
30. A person whose z -score was zero would have a raw score equal to
- a) zero.
 - b) the mean.
 - c) the median.
 - d) none of these.
31. If one desires to give equal weight to several tests when combining scores, he will obtain the best results if he uses
- a) percentages.
 - b) percentile scores.
 - c) raw scores.
 - d) standard scores.
 - e) any of the above--it makes no difference.
32. Is it wise to base a pupil's grade on his achievement in relation to his ability? Why?
- a) Yes, because achievement depends on ability.
 - b) Yes, because it gives every pupil an equal chance to make high marks.
 - c) No, because the ability required in a particular course is difficult to assess properly.
 - d) No, because to do so simply intensifies the competition for high marks.

APPENDIX B

THE INSTRUMENT (FINAL EDITION)

To my colleagues who teach in Khon Kaen Province schools,

I am studying for my doctor's degree at Michigan State University, East Lansing, Michigan. The last requirement is that I do a research project. Because when I have finished I expect to return to the Faculty of Education at Khon Kaen University, and because Khon Kaen is my home city, I am studying whether the teachers in our province need and want help in making and using classroom tests.

The Changwad Education Officer and the Amphur Education Officers are interested in helping in this study, and I am very grateful to them. Your Amphur Education Officer is distributing the questionnaires and will return them to me. But nobody except me will know just how you answered. When you have made your answers, put your questionnaire in the enclosed envelope, seal it, and hand it to your principal, who will start it on its way back to me. Do not sign your name on the envelope. Answer each question as accurately and as honestly as you can. Your responses to this questionnaire will not be identified to you individually, therefore do not worry about being wrong.

If my study is to be acceptable and later become useful, it is important that you take this task seriously and I get back all the questionnaires. It will take about thirty minutes for you to fill out the form.

Thank you very much for your help. The findings will probably make the basis for designing a course or workshop to be offered to all teachers.

Cordially Yours,

Sor Wasna Pravalpruk

Directions:

You need not identify yourself, but please fill in all information on this page. This information is for research purposes only.

Answer each of the questions on this and the following pages as accurately and as honestly as you can.

PART I Please fill in the blanks and check the appropriate categories.

Name of school _____

Amphur _____

What grade are you teaching? _____

What subjects are you teaching? _____

How many teachers are there in your school? _____

How many students are there in your school? (approximate)

Your sex ☐ Male ☐ Female

Level of teacher education:

☐ Cert. and lower

☐ Higher cert. and higher

Years of teaching experience:

☐ Less than 8 years

☐ Eight years and more

Did you attend the Measurement and Evaluation
Workshop program during 1966-1968?

☐ Yes

☐ No

Are you in favor of national testing?

☐ Yes

☐ No

PART II Please mark X in the appropriate columns for each item.

Do you think you need to know more about these tasks?

	Urgent Need	Very Much Needed	Some Need	Little Need	Don't Need
1. Choosing the type of test to use					
2. Determining the length of the test					
3. Deciding what ability should be emphasized in the test					
4. Determining the number of items for each topic					
5. Administering the test					
6. Setting the difficulty level					
7. Arranging the items into the test					
8. Writing good items stems (in multiple-choice item)					
9. Writing reasonable alternatives (in multiple-choice item)					
10. Writing good true-false items					
11. Scoring essay tests					
12. How to measure in the cognitive domain					
13. How to measure in the affective domain					
14. How to measure in the psycho-motor domain					
15. Understanding the major uses and limitations of item analysis data					

	Urgent Need	Very Much Needed	Some Need	Little Need	Don't Need
16. Computing and understanding the item difficulty index					
17. Computing and understanding the item discrimination index					
18. Understanding the difference and relationship between difficulty and discrimination					
19. Finding the validity of a test					
20. Finding the reliability of a test					
21. Revising items					
22. Reporting and writing recommendations to the principal and superintendent					
23. Choosing the type of test scores					
24. Computing and understanding T-scores					
25. Computing and understanding percentiles					
26. Drawing and interpreting profiles					
27. Using norms					
28. Distinguishing between norms and standard					

PART III Mark X in front of the best answer of each item.

1. In evaluating an achievement test for possible use in your school, the most important issue is whether the test
 - a) is based upon a thorough survey of teaching practices over the whole country.
 - b) matches in content the objectives of instruction in the school.
 - c) has sufficiently high predictive validity.
 - d) has sufficiently high reliability.
2. The most important concern in judging how well a test samples a content area is its
 - a) length.
 - b) correlation with a criterion.
 - c) representativeness.
 - d) variety.
3. In constructing a table of specifications for a classroom test, the number of test items allotted to each cell should be
 - a) varied each term, so that students are unable to detect the amount of emphasis to be given to each area in the test.
 - b) approximately equal, so that no particular objective will receive undue emphasis.
 - c) balanced between knowledge and higher level learning outcomes.
 - d) determined by the emphasis to be given during instruction.
4. Which of the following will not be accomplished by constructing a test blueprint?
 - a) improve the validity of the test.
 - b) relate evaluation to teaching.
 - c) help to "balance" the test.
 - d) save construction time.

5. Which type of item format provides greatest content sampling in the shortest possible testing time?
 - a) matching.
 - b) multiple-choice.
 - c) short-answer.
 - d) true-false
6. Multiple-choice items written in the form of questions
 - a) assume complex reading skills.
 - b) require no more detailed alternatives than do incomplete statements.
 - c) have greater clarity than items phrased in incomplete statements.
 - d) lend themselves to easier detection of construction flaws.
7. It is good practice in multiple-choice, single-correct answer test items to
 - a) randomize the order of the answers, putting the correct alternative in each position an approximately equal number of times.
 - b) put the correct answer last or next to last more often than in any other position.
 - c) put the correct answer first more often than in any positions.
 - d) arrange placement of the correct answers according to some predetermined scheme.
8. Which, if any, of the following should characterize the stem of a good multiple-choice test item?
 - a) It should be the subject portion of the proposition the item is intended to test.
 - b) It should express or imply a specific question.
 - c) It need not have any of the foregoing characteristics.
 - d) Both (a) and (b).

9. To make a multiple-choice item easier, we should make the
- a) question more specific, responses more alike.
 - b) question more specific, responses more diverse.
 - c) question more general, responses more alike.
 - d) question more general, responses more diverse.
10. Within a given item format, the items of a test should be arranged by
- a) decreasing level of complexity.
 - b) increasing level of complexity.
 - c) level of reading difficulty.
 - d) a random pattern.
11. Fundamentally, item discrimination indices attempt to indicate
- a) the ability of items to distinguish between persons having different levels of that characteristics which the test measures.
 - b) how well the items can be distinguished from one another within the test.
 - c) the degree to which items from different tests can be discriminated.
 - d) the ability of the item writer to develop items which are readable.
12. If three good students out of four answer a question correctly, what is its index of discrimination?
- a) 1.00
 - b) .75
 - c) .25
 - d) Insufficient information is given to answer the question.

Items 13 through 15 are based on the following data:

A class of 200 students was divided at the median into a high and a low groups. Their responses to five 4 choice multiple-choice items are tabulated below. Correct responses are underlined.

<u>Number of students marking</u>					
<u>Item Number</u>	<u>Group</u>	<u>Response 1.</u>	<u>Response 2.</u>	<u>Response 3.</u>	<u>Response 4.</u>
1	High	<u>80</u>	20	0	0
	Low	<u>45</u>	30	20	5
2	High	0	0	<u>100</u>	0
	Low	0	15	<u>85</u>	0
3	High	<u>35</u>	10	10	45
	Low	<u>55</u>	20	0	25
4	High	20	5	<u>65</u>	10
	Low	20	25	<u>35</u>	20

Use the following KEY to answer each of the next four items:

- KEY: a) Item 1
 b) Item 2
 c) Item 3
 d) Item 4

13. Which item discriminated best between high and low groups? _____
 14. Which item has difficulty index of .50? _____
 15. Which item shows negative discrimination? _____

16. The norms for which type of test need to be revised more often?
- a) Achievement
 - b) Aptitude
 - c) Interest
 - d) Projective
17. On test XYZ pupil A scored at the 80th percentile and pupil B scored at the 40th percentile. What statement may we make relative to their abilities?
- a) Pupil A has twice the measured ability of Pupil B.
 - b) Pupil A's test score will be twice that of Pupil B.
 - c) If Pupil A's score is superior to those of 70 pupils, B's score is superior to those of 35.
 - d) Percentile ranks are meaningless in comparing the scores of A and B.
18. A person whose z-score was zero would have a raw score equal to
- a) zero.
 - b) the mean.
 - c) the mode.
 - d) none of these.
19. If one desires to give equal weight to several tests when combining scores, he will obtain the best results if he uses
- a) percentile scores.
 - b) raw scores.
 - c) standard scores.
 - d) any of the above--it makes no difference.

20. Is it wise to base a pupil's grade on his achievement in relation to his ability? Why?
- a) Yes, because achievement depends on ability.
 - b) Yes, because it gives every pupil an equal chance to make high marks.
 - c) No, because the ability required in a particular course is difficult to assess properly.
 - d) No, because to do so simply intensifies the competition for high marks.

เรียน เพื่อนครูในจังหวัดขอนแก่น

ดิฉันกำลังศึกษาเพื่อรับปริญญาเอกอยู่ที่ **Michigan State University**

ในสหรัฐอเมริกา งานชิ้นสุดท้ายที่จะต้องทำคือปริญญาโท เนื่องจากเมื่อจบการศึกษาแล้ว ดิฉันจะไปทำงานที่มหาวิทยาลัยขอนแก่นในคณะศึกษาศาสตร์ และขอนแก่นเป็นบ้านเกิดของดิฉัน ดิฉันจึงทำการศึกษา เรื่องความต้องการของครูในจังหวัดของเรา เกี่ยวกับการออกข้อสอบและวัดผล นักเรียน

ดิฉันได้ขอความร่วมมือจากท่านศึกษาธิการจังหวัดและอำเภอซึ่งท่านได้ให้ความร่วมมือในโครงการนี้ ขอขอบพระคุณท่านไว้ ณ ที่นี้ด้วย ท่านศึกษาธิการอำเภอและจังหวัดจะเป็นผู้แจกแบบสอบถามนี้ให้ตามโรงเรียนโดยผ่านครูใหญ่ เมื่อท่านตอบคำถามทั้งหมดเสร็จแล้ว กรุณาส่งให้ครูใหญ่หรือศึกษาธิการอำเภอหรือจังหวัดเพื่อจะได้รวบรวมส่งให้ดิฉันต่อไป

แบบสอบถามชุดนี้จะใช้เพื่อการศึกษาและวิจัยของดิฉันเท่านั้น และจะไม่พิจารณาคำตอบของท่านเป็นรายบุคคล ท่านไม่ต้องเขียนชื่อของท่าน เพียงแต่กรอกข้อความต่างๆให้ครบทุกข้อเท่านั้น

แบบสอบถามมี ๓ ตอน ตอนที่ ๓ อาจจะเป็นวิชาการมากไป ซึ่งดิฉันก็ได้พยายามหลีกเลี่ยงให้ให้น้อยที่สุด ขอให้ท่านพยายามตอบให้หมดทุกข้อ

ผลของการศึกษาเรื่องนี้จะมีประโยชน์ต่อไปภายหน้าก็ด้วยความร่วมมือในการตอบแบบสอบถามนี้ทุกข้อทุกตอนตามความนึกคิดของท่านและส่งแบบสอบถามคืนให้ดิฉัน ขอขอบพระคุณในความร่วมมือของท่านมา ณ โอกาสนี้ด้วย

ด้วยความนับถือ

(ส.วาสนา ประवालพฤกษ์)

คำชี้แจง

โปรดตอบคำถามทุกข้อให้ชัดเจน ข้อความต่างๆในแบบสอบถามนี้จะใช้ในการศึกษาวิจัยเท่านั้น

ตอนที่ ๑ กรุณาเติมข้อความในช่องว่าง หรือกา ลงในช่องที่เหมาะสม

อำเภอที่ท่านทำการสอน

ชั้นที่สอน

วิชาที่สอน

จำนวนครูในโรงเรียน

จำนวนนักเรียน (ประมาณ)

เพศของท่าน ☐ ชาย ☐ หญิง

วุฒิกู ☐ ปกศ. หรือต่ำกว่า

☐ ปกศ.สูง ปริญญา หรือสูงกว่า

จำนวนปีที่ท่านสอน ☐ น้อยกว่า ๔ ปี

☐ ๔ ปีและมากกว่า ๔ ปี

ท่านเคยเข้าอบรมการวัดผลบ้างหรือไม่

☐ เคย เมื่อปี

☐ ไม่เคย

ท่านขอให้มีการสอบรวมทั้งประเทศหรือไม่

☐ ชอบ ☐ ไม่ชอบ

ตอนที่ ๒ ท่านอยากจะศึกษาเพิ่มเติมเกี่ยวกับเรื่องต่อไปนี้อย่างไรหรือไม่เพียงใด ถ้าท่านคิดว่าต้องการมากที่สุด ให้กา X ลงในช่องที่ ๑ (ต้องการมากที่สุด) และที่ ๒,๓,๔, รองลงไปจนถึงช่องที่ ๕ คือไม่ต้องการเลย

ต้องการมากที่สุด

ไม่ต้องการเลย

๑. การเลือกประเภทของข้อสอบ เช่น ประนัยหรืออัตนัย
๒. การกำหนดความยาวของข้อสอบ
๓. ข้อสอบควรเน้นสมรรถภาพทางสมองอะไรบ้าง
๔. การกำหนดว่าหัวข้อใดควรจะทำข้อสอบกี่ข้อ
๕. วิธีการคุมสอบ
๖. การกำหนดความยากง่ายของข้อสอบ
๗. การเรียงข้อสอบแต่ละข้อ เป็นข้อสอบทั้งฉบับ
๘. การเขียนตัวคำถามในข้อสอบแบบ เลือกตอบ
๙. การเขียนตัวเลือกในข้อสอบแบบ เลือกตอบ
๑๐. การเขียนข้อสอบแบบถูกผิด
๑๑. การให้คะแนนข้อสอบแบบอัตนัย
๑๒. การออกข้อสอบวัดปัญญา เช่นในวิชาวิทยาศาสตร์
๑๓. การวัดความรู้ลึกหรือทัศนคติของนักเรียน
๑๔. การวัดทักษะ เช่นในวิชาพลศึกษา ทัศนศึกษา
๑๕. ข้อกำหนดกฎเกณฑ์ในการวิเคราะห์ข้อสอบ
๑๖. การคำนวณและความหมายของความยากง่ายของข้อสอบ
๑๗. การคำนวณและความหมายของอำนาจจำแนกของข้อสอบ
๑๘. ความแตกต่างระหว่างความยากง่ายและอำนาจจำแนกของข้อสอบ

[illegible]

๑๙. การหาความเที่ยงตรงของข้อสอบ
๒๐. การหาความเชื่อมั่นของข้อสอบ
๒๑. การปรับปรุงแก้ไขข้อสอบ
๒๒. การกรอกสมุดประจำตัวนักเรียนและรายงานต่อครูใหญ่
๒๓. การเลือกใช้ชนิดของคะแนนเพื่อเสนอผลการสอบ
๒๔. การคำนวณและความหมายของคะแนน T
๒๕. การคำนวณและความหมายของเปอร์เซ็นต์ไทล์ (Percentile)
๒๖. การเปรียบเทียบความสามารถในด้านต่างๆของนักเรียนโดยใช้ตาราง Profile
๒๗. การใช้เกณฑ์ปกติ (Norm)
๒๘. ข้อแตกต่างระหว่างเกณฑ์ปกติและมาตรฐาน (Standard)

[illegible]

ตอนที่ ๓ กรุณาตอบคำถามข้างล่างนี้ตามที่ท่านทราบ พยายามตอบทุกข้อ กา ~~X~~ หน้าคำตอบที่ท่านเห็นว่าถูก

๑. ข้อสอบที่จะวัดความสัมฤทธิ์ผลในโรงเรียนของท่านควรจะมี
 - ก. วัดในสิ่งที่ครูทุกคนเห็นสมควร
 - ข. วัดตามที่กำหนดไว้ในจุดมุ่งหมายของวิชานั้นๆ
 - ค. มีความเที่ยงตรงสูง
 - ง. มีความเชื่อมั่นสูง
๒. ในการพิจารณาว่าข้อสอบครอบคลุมในเนื้อหาวิชาเพียงใดนั้น เราควรที่
 - ก. ความยาวของข้อสอบ
 - ข. ความสัมพันธ์ของข้อสอบกับกฎเกณฑ์ที่ตั้งไว้
 - ค. ความเป็นตัวแทนของข้อสอบต่อวิชานั้น
 - ง. ข้อสอบว่าแต่ละข้อมีความแตกต่างกันในเนื้อหา
๓. ในการกำหนดว่าในแต่ละตอนของวิชาหนึ่งๆควรจะมีข้อสอบกี่ข้อ ควรจะมี
 - ก. เปลี่ยนทุกครั้งที่มีการสอบเพื่อให้นักเรียนรู้ว่าเรื่องนี้จะมีการถามในข้อสอบเท่านี้ข้อ
 - ข. มีเท่าๆกันในแต่ละตอนของเนื้อหาวิชา
 - ค. ควรจะมีวัดเท่าๆกันในด้านความจำและการใช้เหตุผล
 - ง. ควรจะกำหนดโดยครูผู้สอน
๔. การทำการางวิเคราะห์หลักสูตรช่วยครูในด้านต่อไปนี้ ยกเว้นข้อไหน
 - ก. ช่วยให้ข้อสอบมีความเที่ยงตรงยิ่งขึ้น
 - ข. ช่วยให้ความสัมพันธ์ในการเรียนและการสอน
 - ค. ช่วยกำหนดจำนวนข้อในข้อสอบ
 - ง. ช่วยประหยัดเวลา
๕. ข้อสอบชนิดไหนที่ออกได้มากที่สุดโดยใช้เวลาในการตอบน้อยที่สุด
 - ก. แบบจับคู่
 - ข. แบบเลือกตอบ
 - ค. แบบให้เขียนคำตอบสั้นๆ
 - ง. แบบถูกผิด

๖. ในการเขียนข้อสอบแบบเลือกตอบ ใช้ประโยคคำถามเพื่ออะไร
- ให้นักเรียนต้องใช้ความสามารถในการอ่านตีความหมายในระดับสูง
 - ตัวเลือกจะเขียนได้สั้น ไม่ต้องอธิบายมาก
 - ตัวเลือกจะได้กระจายกันไปหลายแง่หลายมุม
 - กำหนดแนวของข้อนั้นได้ชัดเจนว่าต้องการถามอะไร
๗. ในข้อสอบแบบเลือกตอบ การให้ตัวคำตอบถูกอยู่ในข้อ ก.ข.ค.ง. หรือจ. นั้นควรกำหนดอย่างไร
- ให้คำตอบถูกอยู่ในข้อ ก.ข.ค.ง.จ. คละกันไปเท่าๆกันในข้อสอบหนึ่งฉบับ
 - จัดให้คำตอบอยู่ท้ายๆ เช่นข้อ จ. ให้มากกว่าตัวเลือกอื่นๆ
 - จัดให้คำตอบอยู่ต้นๆ เช่นข้อ ก. ให้มากกว่าตัวเลือกอื่นๆ
 - จัดคำตอบถูกให้เรียงกันแบบใดแบบหนึ่งตามหลักการที่กำหนดไว้
๘. ตอนนำในข้อสอบแบบเลือกตอบควรมีลักษณะอย่างไร
- ควรเป็นประธานของประโยคเพื่อให้คำตอบในตอนเลือกเป็นส่วนขยาย
 - ควรเป็นประโยคที่ถามเฉพาะเจาะจง
 - เป็นประโยคแบบใดก็ได้
 - ทั้ง ก. และ ข.
๙. เพื่อจะให้ข้อสอบแบบเลือกตอบง่ายขึ้นกว่าเดิม ควรทำให้
- ตอนนำถามจำเพาะลงไป และตอนเลือกใกล้เคียงกันมากๆ
 - ตอนนำถามจำเพาะลงไป และตอนเลือกต่างกันมากๆ
 - ตอนนำถามให้กว้าง และตอนเลือกใกล้เคียงกันมากๆ
 - ตอนนำถามให้กว้าง และตอนเลือกต่างกันมากๆ
๑๐. การเรียงข้อสอบแต่ละข้อเข้า เป็นข้อสอบทั้งฉบับควร เรียงอย่างไร
- เรียงจากข้อยากไปง่าย
 - เรียงจากข้อง่ายไปยาก
 - เรียงตามลำดับข้อยาว-สั้น
 - เรียงอย่างไรก็ได้

๑๑. อ่านาจจำแนกของข้อสอบคืออะไร

- ก. ความสามารถที่จะแบ่งนักเรียนออกตามระดับความสามารถ
- ข. ความสามารถที่จะแบ่งข้อสอบแต่ละข้อออกจากข้ออื่นๆในข้อสอบทั้งฉบับ
- ค. คำนึงบอกความแตกต่างของข้อสอบแต่ละข้อออกจากข้อสอบหลายๆฉบับ
- ง. ความสามารถของผู้ออกข้อสอบที่จะปรับปรุงข้อสอบแต่ละข้อให้ใช้ได้

๑๒. ในข้อสอบข้อหนึ่งมีนักเรียนเก่ง ๓ ใน ๔ คนตอบข้อนั้นถูก ถามว่าข้อสอบข้อนี้มีอ่านาจจำแนกเท่าไร

- ก. ๑.๐๐
- ข. .๗๕
- ค. .๒๕
- ง. ยังตอบไม่ได้ เพราะต้องทราบอย่างอื่นเพิ่มอีก

ใช้ข้อความข้างล่างนี้ตอบข้อ ๑๓ ถึง ๑๕

ในชั้นหนึ่งมีนักเรียน ๒๐๐ คน แบ่งออกเป็น ๒ กลุ่มคือกลุ่มคะแนนสูงและกลุ่มคะแนนต่ำ กลุ่มละ ๑๐๐ คน ตารางข้างล่างนี้เป็นคำตอบของข้อสอบชนิดเลือกตอบ ๔ ข้อ (คำตอบถูกคือคำตอบที่ขีดเส้นใต้)

ข้อ	กลุ่มนักเรียน	จำนวนนักเรียนที่ตอบแต่ละตัวเลือก			
		ตัวเลือก ก.	ตัวเลือก ข.	ตัวเลือก ค.	ตัวเลือก ง.
๑.	สูง	<u>๘๐</u>	๒๐	๐	๐
	ต่ำ	<u>๔๕</u>	๓๐	๒๐	๕
๒.	สูง	๐	๐	<u>๑๐๐</u>	๐
	ต่ำ	๐	๑๕	<u>๘๕</u>	๐
๓.	สูง	<u>๓๕</u>	๑๐	๑๐	๔๕
	ต่ำ	<u>๕๕</u>	๒๐	๐	๒๕
๔.	สูง	๒๐	๕	<u>๖๕</u>	๑๐
	ต่ำ	๒๐	๒๕	<u>๓๕</u>	๒๐

ใช้ตัวเลือกข้างล่างนี้ตอบข้อ ๑๓ ถึง ๑๕

- ก. ข้อ ๑
- ข. ข้อ ๒
- ค. ข้อ ๓
- ง. ข้อ ๔

- ๑๓. ข้อไหนมีอำนาจจำแนกสูงที่สุด
- ๑๔. ข้อไหนมีระดับความยากง่าย .๕๐
- ๑๕. ข้อไหนมีอำนาจจำแนกเป็นลบ

๑๖. เกณฑ์ปกติ (Norms) ของข้อสอบประเภทไหนที่ต้องปรับปรุงแก้ไขบ่อยที่สุด

- ก. ข้อสอบความสัมฤทธิ์ในการศึกษา
- ข. ข้อสอบวัดความถนัด
- ค. ข้อสอบวัดความสนใจ
- ง. ข้อสอบวัดทัศนคติ

๑๗. ในข้อสอบฉบับหนึ่ง เด็กชายก. ได้คะแนนที่ ๔๐ เปอร์เซนต์ไธล์ (Percentile) เด็กชายข. ได้คะแนนที่ ๔๐ เปอร์เซนต์ไธล์ ถามว่าข้อความข้างล่างนี้ข้อใดถูกต้อง

- ก. ก.มีความสามารถเป็น ๒ เท่าของ ข.
- ข. คะแนนของ ก. เป็น ๒ เท่าของ ข.
- ค. ถ้า ก. ได้คะแนนสูงกว่าคนอื่น ๙๐ คน ข. จะได้คะแนนสูงกว่าคนอื่น ๑๕ คน
- ง. ตำแหน่งเปอร์เซนต์ไธล์ไม่มีความหมายต่อการเปรียบเทียบคะแนนของนักเรียน ๒ คน

๑๘. ถ้านักเรียนคนหนึ่งมีคะแนน z เป็นศูนย์ เขามีคะแนนดิบ (คะแนนที่ได้จากข้อสอบโดยตรง) เท่าไร

- ก. ศูนย์
- ข. เท่ารายเฉลี่ย (Mean)
- ค. เท่าคะแนนที่มีคนได้มากที่สุด (Mode)
- ง. ไม่มีคำตอบถูก

๑๙. ครูคนหนึ่งต้องการรวมคะแนนจากข้อสอบหลายฉบับโดยใช้อัตราส่วนหรือน้ำหนักของข้อสอบทุกฉบับเท่ากัน เขาควรจะใช้คะแนนชนิดใดจึงจะดีที่สุด

ก. เปอร์เซนต์ไทล์ (Percentile)

ข. คะแนนดิบ (Raw Score)

ค. คะแนนมาตรฐาน (Standard Score)

ง. ใช้คะแนนแบบไหนก็ได้ ไม่ทำให้ความหมายแตกต่างกัน

๒๐. ครูควรจะให้คะแนนหรือเกรด Grade นักเรียนโดยดูจากผลงานที่นักเรียนทำเทียบกับความสามารถหรือความถนัดของนักเรียนคนนั้นหรือไม่

ก. ควร เพราะผลงานขึ้นอยู่กับความสามารถ

ข. ควร เพราะทำให้นักเรียนมีโอกาสที่จะได้รับคะแนนสูงเท่าๆกัน

ค. ไม่ควร เพราะความสามารถในวิชาหนึ่งๆมีองค์ประกอบที่ไม่แน่นอน

ง. ไม่ควร เพราะทำให้เกิดการแข่งขันที่จะได้คะแนนสูง

APPENDIX C

APPENDIX C

PAIRWISE PROFILE ANALYSIS ON ACTUAL NEEDS

Sources	df	MS	F	P less than
All Groups				
PL-RE	1	130.46	128.05	.0001
IE-RE	1	178.45	168.16	.0001
IA-RE	1	30.18	25.83	.0001
Location				
PL-RE	1	0.35	0.35	.557
IE-RE	1	3.31	3.11	.079
IA-RE	1	1.70	1.46	.229
Teacher Education				
PL-RE	1	0.04	0.04	.845
IE-RE	1	1.06	1.00	.319
IA-RE	1	1.50	1.28	.258
Teaching Experience				
PL-RE	1	5.44	5.44	.022
IE-RE	1	3.31	3.11	.079
IA-RE	1	8.79	7.52	.007
Location x Education				
PL-RE	1	0.98	0.96	.328
IE-RE	1	0.76	0.71	.399
IA-RE	1	1.14	0.97	.324

PARWISE PROFILE ANALYSIS ON ACTUAL NEEDS
continued

Sources	df	MS	F	P less than
Location x Experience				
PL-RE	1	1.50	1.47	.226
IE-RE	1	0.01	0.01	.939
IA-RE	1	0.69	0.59	.443
Education x Experience				
PL-RE	1	7.44	7.30	.007
IE-RE	1	6.80	6.41	.012
IA-RE	1	11.82	10.12	.002
Location x Education x Experience				
PL-RE	1	11.29	11.07	.001
IE-RE	1	1.06	1.00	.319
IA-RE	1	0.98	0.84	.361
Within				
PL-RE	312	1.019		
IE-RE	312	1.061		
IA-RE	312	1.169		

APPENDIX D

APPENDIX D

MULTIVARIATE ANALYSIS OF VARIANCE ON FOUR SUBSCALES ON ACTUAL NEEDS

Sources	df	MS	F	P less than
Location	4 & 309		2.258	.063
Teacher Education	4 & 309		2.554	.039
PL	1	0.80	0.73	.393
IE	1	6.90	5.44	.020
IA	1	8.45	5.84	.016
RE	1	1.38	1.27	.261
Experience	4 & 309		2.549	.040
PL	1	0.05	0.05	.831
IE	1	0.25	0.20	.655
IA	1	1.25	0.86	.353
RE	1	9.45	8.69	.004
Location x Teacher Education	4 & 309		1.219	.303
Location x Experience	4 & 309		1.504	.201
Education x Experience	4 & 309		3.065	.017
PL	1	1.25	1.15	.286
IE	1	0.90	0.71	.400
IA	1	4.51	3.12	.078
RE	1	7.50	6.89	.009

MULTIVARIATE ANALYSIS OF VARIANCE
ON FOUR SUBSCALES ON ACTUAL NEEDS

continued

Sources	df	MS	F	P less than
Location x Teacher Education x Experience	4 & 309		3.618	.007
PL	1	3.20	2.93	.088
IE	1	2.28	1.80	.181
IA	1	2.45	1.67	.194
RE	1	8.78	8.07	.005
Within				
PL	312	1.092		
IE	312	1.268		
IA	312	1.447		
RE	312	1.088		

BIBLIOGRAPHY

BIBLIOGRAPHY

- Ahmann, J.S., and Glock, M.D. Evaluating Pupil Growth. Boston, Massachusetts: Allyn and Bacon, Inc., 1971.
- Allen, Margaret E. "Status of Measurement Courses for Undergraduates in Teacher-Training Institutions." Thirteenth Yearbook National Council on Measurements Used in Education, 1956: 67-73.
- Bell, R.D. "The Risky Business of Classroom Testing." Business Education Forum. 27 (October 1970): 8-10.
- Black, H. They Shall Not Pass. New York: William Morrow & Co., Inc., 1963.
- Bowman, A. "In-Service Training of Teachers in Measurement." The 18th Yearbook of the National Council on Measurement in Education. 1961: 43-49.
- Brimm, L. and Tollett, D.J. "How do Teachers Feel About In-Service Education?" Educational Leadership. 31 (March 1974): 521-525.
- Durost, N. "Problems in In-Service Training of Teachers in the Use of Measurement and Evaluation Techniques." The 16th Yearbook of the National Council on Measurement in Education. 1959: 31-33.
- Dyer, S. "Testing's Nine Major Pitfalls." Senior Scholastic. 80 (March 14, 1962): 21T-23T.
- Ebel, R.L. (Ed.) Encyclopedia of Educational Research. 4th ed., 1969.
- Ebel, R.L. Essentials of Educational Measurement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.
- Ebel, R.L. Essentials of Educational Measurement: Instructor's Manual. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.
- Ebel, R.L. "Improving the Competence of Teachers in Educational Measurement." Clearing House. 36 (October 1961): 67-71.
- Ebel, R.L. "Measurement and the Teacher." Educational Leadership. 20 (October 1962): 20-24.
- Ebel, R.L. "The Social Consequences of Educational Testing." School and Society. 92 (1964): 331-334.
- Findley, W.G. (Ed.) "The Impact and Improvement of School Testing Program." (Article by Elizabeth Hagen and Lucille Lindberg) The 62nd Yearbook of the National Society for the Study of Education Part II, 1963: 232-253.

- Findley, W.G., and Bryan, Miriam M. "Tests Like Handguns..." (From Education Daily, January 20, 1972). Phi Delta Kappan 53 (March 1972): 442.
- Gerberich, J.R. "The Development of Educational Testing." Theory Into Practice. 2 (October, 1963): 184-191.
- Goslin, D.A. Teachers and Testing. Russell Sage Foundation, 1967.
- Grobman, Hulda "Classroom Testing and the Biology Teacher: An Annotated Bibliography." American Biology Teacher. 29 (April, 1967): 282-285.
- Grobman, Hulda "Classroom Testing in Biology: An Annotated Bibliography (II)." American Biology Teacher. 33 (February 1971): 86-90.
- Heenan, D.K. and Chaval Paeratakul Education in Thailand: Evaluation of Instruction in Thailand. East Lansing, Michigan State University, 1966.
- Holt, J. "I Oppose Testing, Marking and Grading." Today's Education. 60 (March 1971): 28-31.
- Jessel, J.C. "Focus on Issues: Should We Declare a Moratorium on Mental Testing." Contemporary Education. 44 (November 1972): 127.
- LaSage, W. "The Hassle Over Standardized Testing." Instructor. 82 (March 1973): 45-56.
- Link, F.R. "Teach-Made Tests." National Education Association Journal. 52 (October 1963): 23-25.
- McMorris, R.F. et al "Effects of Violating Item Construction Principles." Journal of Educational Measurement. 9 (Winter 1972): 287-295.
- McSwain, E.T. "The Role of the College Faculty in the Teacher Education Program." The 16th Yearbook of the National Council on Measurement in Education. 1959: 34-38.
- Mehrens, W.A. "The Consequences of Misusing Test Results." National Elementary Principal. 47 (September 1967): 62-64.
- Mehrens, W.A. and Lehmann, I.J. Measurement and Evaluation in Education and Psychology. New York: Holt, Rinehart and Winston, Inc., 1973.
- Mehrens, W.A. and Lehmann, I.J. Measurement and Evaluation in Education and Psychology: Instructor's Manual. New York: Holt, Rinehart and Winston, Inc., 1973.
- Noll, V.H. "Problems in the Pre-Service Preparation of Teachers in Measurement." The 18th Yearbook of the National Council on Measurement in Education. 1961: 35-42.

- Noll, V.H. "Requirements in Educational Measurement for Prospective Teachers." School and Society. 82 (September 17, 1955): 88-90.
- Roeder, H.H. "Are Today's Teachers Prepared to Use Tests?" Peabody Journal of Education. 49 (April 1972): 239-240.
- Rose, H.A. and Elton, C.F. "Testing, on the Bias." Journal of College Student Personnel. 12 (September 1971): 362-364.
- Rosner, B. "Substance of Training in Measurement and Evaluation in Pre-Service Preparation of Teachers." The 18th Yearbook of the National Council on Measurement in Education. 1961: 51-54.
- Schutz, R.E. "The Role of Measurement in Education: Servant, Soulmate, Stoolpigeon, Statesman, Scapegoat, All of the Above, and/or None of the Above." Journal of Educational Measurement. 8 (Fall 1971): 141-146.
- Schwartz, A.A., and Tiedeman, D. Evaluating Student Progress in the Secondary School. New York: Longmans, 1957.
- Sharp, Evelyn W. "How Teachers See Testing." Educational Leadership 24 (November 1966): 141-143.
- Stevens, C.A. "Measurement in the Teaching-Learning Situation." The 16th Yearbook of the National Council on Measurement in Education. 1959: 39-42.
- Stevens, Margaret "The Role of the Classroom Teacher in School Testing Programs." The 16th Yearbook of the National Council on Measurement in Education. 1959: 43-46.
- Thorndike, R.L. and Hagen, Elizabeth Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, Inc., 1969.
- Trump, J.L. "What's Wrong with Testing." Theory Into Practice. 2 (1963): 235-242.
- Tylor, Jean L. "Of Tools and Fools." Phi Delta Kappan. 55 (December 1973): 285-286.
- Wronski, S.P. and Kaw Sawasdi-Panich Education in Thailand: Secondary Education, Manpower and Education Planning in Thailand. East Lansing: Michigan State University, 1966.



MICHIGAN STATE UNIV. LIBRARIES



31293100191877