

## VOLUME 9 · CENTER FOR EDUCATIONAL MEASUREMENT, INC. • 1998-1999

**Contemporary Psychological Theories of Intelligence** and Their Implications for Educational Assessment ALLAN B. I. BERNARDO

Techniques of Assessing, Types, and Taxonomy of Measures of Thinking Skills **RICHARD DLC. GONZALES** 

The Learning Environment and Supervisory Practices of a Private Secondary School ZULKIFLI A. MANAF and RAHIMAH HAJI AHMAD

**Factors Related to Student Ratings on Teacher Effectiveness** SIMON LLOYD D. RESTUBOG



CENTER FOR EDUCATIONAL MEASUREMENT, INC.

# Philippine Journal of Educational Measurement

#### **VOLUME 9**

1998-1999

The *Philippine Journal of Educational Measurement* is published annually by the Center for Educational Measurement, Inc., and contains empirical researches and non-empirical reports such as theoretical notes, specialized reviews, and commentaries on issues related to education i.e., curriculum, instruction, testing, assessment, and evaluation.

Editor

Lenore Ll. Decenteceo, PhD

Assistant Editor

Venia Fe E. Pepito

Staff

Myrna A. Diaz Dennis M. Marciano Joy J. Macalintal ISSIN: 1656-8575 @ 2105 181 Isome: 1989-1990

## ORDERING INFORMATION

Subscriptions are on a per-volume basis at a rate of P200.00. Discounts are available for quantity purchases. All orders must be payable by postal money order to the Center for Educational Measurement, Inc. Business correspondence, including orders and remittances relating to subscriptions, should be addressed to PJEM Subscriptions, CEM Publications, 24th Floor Cityland Pasong Tamo Tower, 2210 Chino Roces Ave., Makati City. Subscribers may arrange for first class or airmail delivery of publications by paying the additional postage (rates on request).

## INSTRUCTION TO CONTRIBUTORS

Address all editorial correspondence, including manuscripts for submission, to the Editor, Philippine Journal of Educational Measurement, CEM Research Unit, 24th Floor Cityland Pasong Tamo Tower, 2210 Chino Roces Ave., Makati City. Unsolicited manuscripts are welcomed. Authors are responsible for the authenticity of all quoted matter and bibliographic citations. The publisher and the Editor assume no responsibility for statements made or opinions expressed by contributors in articles printed in the Philippine Journal of Educational Measurement, and likewise, do not necessarily reflect official CEM policy.

Manuscripts are carefully reviewed by the Editor and one or more Editorial Assistants. Proofs will be sent to author if there is sufficient time to do so. They should be corrected and returned to the Editor within seven days. Authors may be requested to make major revisions prior to acceptance. The Editor reserves the right to make editorial changes of nonsubstantive nature.

## COPYRIGHT AND PERMISSION

Papers accepted become the copyright of the Journal, unless otherwise specifically agreed, and they may not be reproduced without the permission of The Philippine Journal of Educational Measurement, and the publisher.

# TABLE OF CONTENTS

Contemporary Psychological Theories of Intelligence	
and Their Implications for Educational Assessment	
ALLAN B. I. BERNARDO	1
Techniques of Assessing, Types, and Taxonomy	
of Measures of Thinking Skills	
RICHARD DLC. GONZALES	13
The Learning Environment and Supervisory Practices	
of a Private Secondary School	
ZULKIFLI A. MANAF and RAHIMAH HAJI AHMAD	25
Factors Related to Student Ratings on Teacher Effectiveness	
SIMON LLOYD D. RESTUBOG	40

and the second second

## CONTEMPORARY PSYCHOLOGICAL THEORIES OF INTELLIGENCE AND THEIR IMPLICATIONS FOR EDUCATIONAL ASSESSMENT

Allan B. I. Bernardo De La Salle University Philippines

This paper first reviews some recent psychological theories of human intelligence, particularly those that derive from the cognitive science perspective on mental functioning. Gardner's Multiple Intelligence Theory, Sternberg's Triarchic Theory and Mental Self-Government Theory, and Glaser's Cognitive Proficiency View were briefly summarized and the major themes underlying these theories were noted. The implications of these theories for education assessment were then discussed. In particular, the discussion focused on issues related to (a) what should be measured by tests for educational assessment, (b) generalizability of test performance, (c) the meaning of norm-referenced test scores, (d) aptitude vs. achievement, (e) the use of more complex and authentic measures, and (f) the purposes for educational assessment.

Educational assessment has had a long alliance with theories of human intelligence. This alliance most likely started during the time Binet was asked by the Paris school board to develop standardized measures that would predict academic success of French students. This alliance continues to be scrutinized and developed to this day.

In the past two or three decades, psychological theories of human intelligence have undergone many dramatic changes. From the established psychometric theories of intelligence that defined human abilities in ters of an empiricallydefined set of factors, psychological theories of intelligence now present a much more complicated picture of the human intellect. The changes have been mainly influenced by the resurgence of cognitive psychology and cognitive science. The cognitive sciences have developed an understanding of human mental functioning as the processing and representation of information. The cognitive science perspective of intellectual functioning and its development gave rise to new ways of thinking about "intelligence" or exceptional intellectual functioning. These new ways of thinking about intelligence manifests in new explicit theories of intelligence and expertise.

In this paper, I will give a brief overview of a few important themes that dominate current psychological theorizing about intelligence and expertise. I will develop these themes by referring to the works of three important cognitive

Philippine Journal of Educational Measurement • Volume 10

theorists Robert Glaser, Howard Gardner, and Robert Sternberg, who represent different points in the range of cognitive theorizing. I will then consider the implications of these themes for educational assessment. In particular, I will consider issues related to (a) defining what educational assessment tools should be measuring (b) generalizability of test performance, (c) the meaning of normreferenced test scores, (d) aptitude vs. achievement, and (e) authenticity and appropriateness of tests.

## Information-Processing Theories of Intelligence

Glaser's indicators of expertise. Glaser (1986) views intelligence as "proficiency or competence in intellectual cognitive performance". His view of intelligence stems from the body of research regarding expert performance and the development of expertise. He draws his ideas from the works of many different cognitive psychologist that reveal the indicators of expertise. According to Glaser (1991) a primary indicator of expertise is the organization of information into coherent, interconnected chunks of knowledge. Knowledge is highly structured, and in effect, becomes accessible in larger units. In contrast, a novice's knowledge is "spotty, consisting of isolated definitions and superficial understanding" (Glaser, 1991, p. 27) and is often limited in accessibility and applicability. The expert's organized chunks of knowledge are referred to as schemata, and these schemata are very useful for perceiving large patterns of information meaningfully, and for relating bits of information to problem goals, solutions, and conditions for action. Also as consequence of this schematization, the expert's knowledge of different operations are proceduralized or automatized, which means they can be executed rapidly, smoothly, and without error.

Related to the development of organized knowledge, experts approach problems in a principled manner. They recognized the underlying principles that determine the problem structure and can design or retrace solutions according to these principles that determine the problem structure and design or retrieve solutions according to these principles. This principled problem solving is usually not found among novices who tend to approach problems on the basis of specific superficial features of a problem task.

Finally, experts also developed very effective skills for regulating and monitoring their problem solving performance. Not only is expert problem solving principled and planful, expert problem solving is also characterized by rapid and regular checks on the appropriateness and effectiveness of the problem-solving strategy. Because of the availability of large amounts of organized knowledge about the problem domain, experts are also better at assessing the relevance of their knowledge, at apportioning time for the different parts of the problem-solving task, at raising questions

about the task, and at predicting the outcomes of different possible solutions.

As these characteristics of expertise depend on the development of knowledge schemata in a domain, expertise or proficiency is domain specific (Glaser, 1984, 1986). Expertise in one domain does not guarantee proficiency in other domans. The domain-specific nature of expertise also stems from the fact that proficiency and expertise in a domain is shaped by a large range of factors which include the specific characteristics and requirements of the problem tasks in the domain.

Gardner's theory of multiple intelligences. Gardner identified seven intelligences: (1) linguistic, (2) logico-mathematical, (3) spatial, (4) musical, (5) bodily-kinesthetic, (6) intrapersonal, and (7) interpersonal. The first two and to some extent the third intelligence, are usually included in traditional definitions of intelligence. However, the musical and bodily-kinesthetic intelligences have traditionally been disregarded as examples of intelligence (e.g., outstanding composers, musicians, athletes, and dancers are not ordinarily thought of as being intelligent). Similarly, the intrapersonal (self-awareness and the ability to access and discriminate among one's feelings) and interpersonal (social adroitness, ability to distinguish among other individuals) intelligences have been thought of as social competencies rather than exemplars of intelligence.

According to Gardner, what defines all these seven intelligences as intelligences is a range of properties that all characterize exceptional performance. All intelligences can be potentially isolated by brain damage. This characteristic reveals that Gardner assumes some degree of modularity in the underlying neurobiological support structure for the intelligences. Each intelligence also has an identifiable core of operations which is distinct from the other intelligences. Each intelligence also has a distinctive developmental history and a definable expert level of performance, themes that reflect Glaser's view about developing specific proficiencies or expertise. All seven intelligences also have an evolutionary plausibility; this theme indicates that Gardner views the intelligences as having an adaptive function in the development of the human species. The differentiation among the intelligences also has support from experimental and psychometric researches. Finally, each of the seven intelligences is susceptible to encoding in a symbol system.

As there are many educational programs in the Philippines that now claim to adopt Gardner's multiple intelligence theory, I wish to underscore that Gardner's theory did not only specify a wider range of domains of intelligent functioning, more importantly Gardner's theory specified important features of the nature and development of these intelligences.

Philippine Journal of Educational Measurement • Volume 10

Sternberg's Triarchic and Mental Self-Government Theory. Sternberg (1985, 1988b) proposed a triarchic theory of intelligence that specifies three subtheories of intelligence, each of which addresses a different aspect of intellectual functioning. These three subtheories are the componential, experiential, and contextual subtheories.

The componential subtheory specifies three sets of information-processing operations or components that serve as the mechanism for intelligent human behaviors. The first set of operations are called the metacomponents and are the cognitive processes involved in planning, monitoring, and decision making during the performance of intellectual tasks. Examples of a matacomponent of intelligence are the process of deciding what problem needs to be solved, organizing the information about the problem, and monitoring one's own solution. The second set of operations in the componential subtheory are the performance components. These are the various processes that are used for the execution of the specific strategies for performing an intellectual task. Examples of performance components are the process of encoding the information about the task, of comparing different problem information, and of responding. The final set of operations in the componential subtheory are the knowledge acquisition components which are used for retrieving and generating new information for doing an intellectual task. Examples of knowledge acquisition components are selective encoding, selective combination, and selective comparison.

The second subtheory of the triarchic theory is the experiential subtheory. The experiential sutheory has two facets, insightful or nonentreched thinking and automated thinking. Insightful thinking refers to the ability to deal with novel tasks. This ability may involve the generation of creative solutions for problems or the creative application of old solutions to new problems. Automated thinking refers to the ability to execute an intellectual function in a fast, smooth, and non-conscious manner. The subtheory is "experiential because it takes into consideration differences in performance that should be related to the amount of experience in a domain. In domains where one has extensive experience, speed and efficiency are the signs of intelligence; but in novel domain, creativity and insight are the hallmark of intelligence.

The contextual subtheory of the triarchic theory addresses how intelligent behaviors relate to the external world. According to Sternberg, intelligence in context consists of the "purposive adaptation to, shaping of, and selection of real-world environments relevant to one's life". Hence, he argues that the behaviors are intelligent in so far as they allow individuals to adapt successfully to the real-world environment. This adaptation does not simply involve adjusting to the environment by way of altering one's repertoire of skills, but also shaping the environment or even deciding to select another environment altogether. Aside from his triarchic theory of intelligence, Sternberg also proposed a theory of intellectual styles which basically advocates that there are different approaches that people may use to engage intellectual tasks. According to him, intelligence involves the ability to govern oneself in the use of the various approaches in ways that will be responsive to one's needs and the requirements of one's environment (Sternberg 1986, 1988a). Hence, he calls this view the mental self-government theory, and he makes analogies between different features and functions of government on the one hand, and the styles of intellectual functioning on the other.

For example, he distinguishes among the legislative, executive, and judicial functions of thought processes. The legislative intellectual functions involve the creation, formulation, and planning of ideas, strategies, and the like. The executive functions involve carrying out the plans formulated by the legislative functions. The judicial functions monitor and provide feedback about the executive functions and its effects after completion. Different tasks will require different combinations of these functions at different points in time. Intelligence involves the ability to manage and engage all appropriate functions throughout the intellectual task.

Sternberg also distinguishes among different forms of problems: monarchic, oligarchic, hierarchic, and anarchic. Monarchic problems specify a single goal, while oligarchic ones require the fulfillment of multiple and equally important goals. Hierarchical problems also require the fulfillment of multiple goals, but the goals have different priorities or valences. Anarchic problems require a clear breakaway from established solutions. According to Sternberg, intellectual functioning may also be global, that is, addressing general and relatively higher levels of abstraction; or it may be local, addressing minutiae involved in conception and implementation of problem solutions. Intellectual functioning may also be conservative, where one tends to solve problems using existing, tried and tested solutions; or it may be progressive, where one attempts to change or extend existing solutions.

The point that Sternberg wished to emphasize is that intelligent functioning cannot be defined by a single criteria. The requirements of the environment and the goals of the individual can be very diverse, hence, intelligent functioning must be defined in terms of a diversity of criteria. In this regard, intelligence may be conceived of as efficacy in mental self-government.

### Summary of Important Themes and Issues

One of the issues confronted by contemporary theories of intelligence is domain-specificity vs. generality of the intellectual functions. Among the

Philippine Journal of Educational Measurement • Volume 10

three theorists, Sternberg takes the side of generality, as he advocates that the different components and features of intelligence are generalizable skills that apply to tasks, regardless of the specific nature or requirements of the task. Gardner assumes specificity of intellectual functioning by virtue of defining seven distinct abilities. Glaser takes the extreme position on domain-specificity in arguing that proficiency develops in relation to the specific characteristic and requirements of the tasks within a domain. However, Glaser and Gardner both assume some common properties of development within each of the domains of intelligence.

Related to the issue of domain-specificity is the issue relating to the important role of knowledge. Glaser's theory emphasizes the role of knowledge \_ its coherence and usefulness\_ in proficient performance, so much so that the hallmark of proficiency is often described in terms of how one's knowledge is organized. Gardner also recognizes the importance of coherent knowledge as part of the development of history of the intelligences. On the other hand, Sternberg does not emphasize knowledge structures at all. All the features of intelligence are related to processes rather than to knowledge. (In some of his works, Sternberg explicitly criticizes theories that emphasize knowledge.) Hence, Sternberg conceives of intelligence as consisting of more abstract abilities than the other theories.

Another issue addressed in current theories of intelligence relates to the nature and **origins of intelligence**. Glaser's emphasis on developing proficiencies seems to argue that intelligent function is something that needs to develop within the domain as one gains more experience in the domain. This view seems to minimize the significance of possible innate differences in ability. Gardner's theory recognizes the biological and evolutionary foundations of the intelligences, and seems to give more weight to innate differences in abilities in one of the intelligences. Although Gardner also argues that all people have all the seven intelligences and that people just differ in terms of how well each is developed. Sternberg's theories seem to be concerned with the origins of intelligence. He does speak of the adaptive character of intelligence, which would suggest that intelligent functioning develops as required in one's requirement. But Sternberg is mostly silent on the origins issue as his focus is on understanding the function and process of intelligent behaviors.

Another issue relates to defining the *scope of intelligence*. Sternberg takes the most traditional view that limits intelligence within the domain of higher cognitive functions, although Sternberg does go beyond traditional definitions of intelligence as applying to a wider range of cognitive skills. Gardner explicitly crosses the boundaries of traditional definitions of intelligence by specifying domains like musical, bodily-kinesthetic, and personal intelligences. Like Sternberg, Glaser had tended to focus on the more cognitive aspects of perfor-

Philippine Journal of Educational Measurement • Volume 10

mance. Work by other researchers that are similar to Glaser's tend to focus on the cognitive aspects of other domains, like the role of problem-perception and strategy selection in domains like sports, judicial decision making and so on.

All theories place importance on *principled problem-solving* and on the metacognitive or control aspects of intellectual functioning. It seems that most theorists of intelligence now think that aspects that are more related to lower or more basic levels of cognition are not as important in characterizing intelligence. These aspects that are currently no longer emphasized include some of th former standards of intelligence testing like memory capacity, computational skills, speed of processing, the accumulation of decontextualized facts, among others.

#### Implications for Educational Assessment

What are the implications of these more contemporary perspectives about intelligence on how we think about educational assessment? There are many points in the different theories that bear on either the theoretical assumptions underlying specific aspects of educational assessment, the realities of educational assessment practices and the issues facing theorists and practitioners of educational assessment. In the following sections I will discuss a few issues and themes on which Glaser's, Gardner's, and Sternberg's theories of intelligence somehow touch. The themes selected for discussion were suggested by Herman's (1991) discussions on related issues.

On what should be measured. One of the first things that we can think about is whether our standardized measures of achievement are measuring the important indicators of learning. At the broadest level, we should think about whether the tests we are using and developing adequately tap the problemsolving skills, knowledge organization, and application in context that are the hallmarks of intelligent functioning according to many theories. Or are the tests we use still tapping isolated bits of fact that are actually not good indicators of intellectual development within a domain?

Many test developers have moved towards having items that tap the higher levels of thinking, reasoning, and problem solving in the different domains. However, it is difficult to have such types of items if one is constrained to use multiple-choice items. We know that it is quite difficult to write multiple items that can tap the higher levels of thinking and problem solving. On the other hand, the items that are easy to make tend to tap the lower levels of cognition. Given that decisions regarding which items are retained in a standardized test is determined empirically (based on empirical coherence), the likelihood that items that tap lower-level cognition remains quite high. The process of relying on empirical coherence to determine which items remain in a standardized test

Philippine Journal of Educational Measurement • Volume 10

also creates problems on the theoretical content validity of such tests. There is no guarantee that the items in the test will reflect a principled understanding of a domain and of what should be measured in the domain. Clearly, we will need to exert greater vigilance at developing tests and test items that assess higher level thinking based on a principled understanding of what higher level thinking in the domain is.

On the generalizability of test performance. Even when one is able to adequately choose items that represent the complexity of one domain, one can wonder whether performance in the domain predicts performance in any other domain. The whole debate on domain-specificity and generality puts into question whatever certitude we may have about the generalizability of performance on specific tests. Even tests on so-called abstract items, may actually be tapping a very specialized domain which is the solving of such abstract problems with such items. There is no guarantee that performance with such abstract reasoning problems will predict reasoning in problems in science, history, or mathematics. Even if one attempts to measure achievement and thinking within a specific areas, one will need to worry about generalizability within that specific area. Does scientific thinking and problem solving in the biology predict scientific thinking and problem solving in physics? Does mathematical thinking in geometry predict mathematical thinking in calculus?

One can always say that the specificity or generalizability of test performance can be determined empirically, that is one can determine whether there are robust correlations among test of specific abilities. However, given the constraints regarding how items are developed and the problems on content validity discussed in the previous section, a robust correlation might not necessarily be indicative of an actual relationship beteen performance in the two domains, rather it may just indicate that the use of similar test-taking skills, or the type of thinking skills required in the different sets of items were similar.

On the meaning of norm-referenced tests. Another important question that can be raise in light of the new theories of intelligence is : What is the meaning of a test result? What is the meaning of a test result in comparison to a standardized norm? We often assume that a test score presented relative to a norm has substantive, criterion-referenced meaning relative to what students know and not know, and can and cannot do. But given the ideas regarding the nature of intelligent functioning and its development expressed in the different theories, the above assumption might actually be a misconception. We cannot assume that two students who get identical test scores in a standardized exam have similar sets or proficiencies, that is, knowledge about concepts and procedures, thinking skills, and so on. The problem is even worse with standardized tests that sample different areas of competency. The composite performance of students in such standardized test is inadequate to predict the performance of a student in any specific domain or much less across domains.

So what is that we know about students based on standardized tests? Herman put her answer in this way, "Generally they simply tell us that students who know a lot in a general area know a lot relative to their peers." But we cannot possibly have anything definite to say about what it is that the student actually knows and does not know.

On the difference between aptitude and achievement tests. One of the first things students of measurement are taught is the difference between aptitude and achievement tests. Aptitude test is supposed to provide a measure of potential or of intrinsic ability. This potential or ability is assumed to be a stable trait, immutable across time and domains of application. On the other hand, achievement test indicates learning, which changes depending in the particular learning experiences one has had. However, given what we know now about how intelligent functions develop we can now question the usefulness of such distinction. In particular it is now difficult to make sense exactly what aptitude tests measure, given concerns about domain specificity, the role of knowledge in development, the particular track of domain for each domain functioning, and other pertinent propositions of Glaser and Gardner. Even the components of intelligence defined by Sternberg, which off-hand seem like generalizable aptitudes, have actually been shown to be malleable over time (see e.g., Brown, Bransford, Ferrara, & Campione, 1983; Weinstein & Meyer, 1991). It seems difficult to determine at any particular point in time whether performance is an indicator of intrinsic abilities or potentials, or is already a product of learning experiences.

On developing more complex assessment tasks. Most of the implications that have been discussed so far raise problems about standardized educational assessment. However, there are also clear directions for improvement being indicated by the recent theories of intelligence. In particular, we can draw from the current understanding of intelligent functioning to try to develop measures of learning and proficient intellectual functioning. In other words, we now seem to know more about how the mind works and what characterizes exceptional mental performance. Moreover, we know what the mind is not, how it does not work, and what does not characterize exceptional mental performance. So there are many guides available for developing more appropriate measures for educational assessment.

One of the clear directions being pointed by the theories we discussed is the need to develop more complex and authentic means of assessing individual learning in schools. An example of a testing program developed based on the recent theories of intellectual functioning is the testing project of Baker,

Freeman, and Clayton (1991). Without describing their program in great detail, some interesting aspects of their project are as follows. First, the test measure focues on important and teachable learning processes, instead of lower level cognitive skills and knowledge. For this purpose, they advocate the use of criterion analysis of these items by experts, rather than using empirical coherence as the determinant of item retention. Second, the measure was sensitive to the impact of prior knowledge, context and task requirements, as these interact with new learned concepts and procedures. Hence, they advocate the use of multiple measures of student learning, where each measure can be statistically and conceptually connected to other indicators of performance. Of course, their approach does not preclude the use of standardized measures as one of the indicators to which other measures can be linked. The point they emphasize is that in order to get a fuller sense of what a student truly knows, the assessment instrument has to be "cognitively-sensitive" which will most likely require that the use of a complex variety of different measures, rather than one standardized instrument.

On defining the purposes of assessment. Probably the most important point of reflection for now should be what our purposes should be for educational assessment. The proposal to have more complex and varied test instruments might not be sensible if one is thinking in terms of regional, national, or international assessment programs. For such large-scale programs, feasibility constraints will require that measures stick to multiple- choice items. Indeed, results from standardized achievement tests can give gross indicators of learning that might be useful for making policy intended to cover a wider range of individuals, particularly in deciding which programs need more support or intervention.

However, for purposes of selecting students for admission, or designing more specific programs of classroom instruction, or for designing learning programs for children who have learning difficulties, or for designing specialized curricula, one might need to rely on assessment measures that give a fuller picture of the individual student's knowledge of procedures, concepts, and how they are related to each other. Currently, the assessment tools used to establish national norms, the instruments used to make cross-national comparisons, and the measures used to evaluate individual learning or ability are all similar in form and substance. What we ought to reflect on is the possibility of using different sets of measures for specific levels of assessment purposes. Maybe, pursuant to the larger imperative of making nationwide assessment of learning across different educational programs, more conventional assessment measures can be used, provided that there is some assurance that these measures actually assess what we intend to assess. Maybe, if the most important goal is to understand what and how individual students learn, we ought to consider using a complex set of more authentic assessment tasks.

#### Summary and Conclusion

To briefly summarize, the paper started by describing the theories of intelligence by three leading cognitive psychologists. The important themes and issues relevant to these theories were then summarized. After which, the different implications of these theories and themes for education assessment were considered. I did not make any specific prescriptions for educational assessment practitioners; rather I propose reflection on several aspects of the educational assessment enterprise that could be modified and improved. Clearly, there is a need to rethink some of our basic assumptions and the practices entailed by these assumptions. There is much room for improvement.

As Glaser (1991) so wisely noted, "The assessment systems we derive depend intimately on our knowledge of how humans learn and acquire knowledge and skill; research and development [of assessment systems] take their cues from findings about the nature of human performance....A scientific base for instruction and its outcomes cannot prosper if we have only minimal understanding of the characteristics of acquired performance."

Therefore, I would like to pose a challenge to the educational assessmen practitioners here today to embrace the task of trying to understand how people develop their intellectual functions based on the emerging and growing perspectives on human intelligence and cognition. Let us repudiate unscientific and outdated notions of learning and intelligence. Let us acknowledge the insights afforded by the extensive range of scientific researches and theories in the area, and the future insights that will be borne by future researches and theories. Let us not rely completely on a theoretical, empirical methods of developing measures for educational assessment. Scientific knowledge on learning and intelligence has blossomed in the past 20 years, we should do our best to ensure that this knowledge benefits the task of educational assessment.

#### References

- Baker, E. L., Freeman, M., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M. C.Wittrock & E.L. Baker (Eds.), *Testing and cognition*. Englewood Cliffs, NJ: Prentice Hall.
- Brown, A.L., Bransford, J. F., Ferrara, R. R., & Campione, J. (1983). Learning, remembering, and understanding. In J. Flavell & E, Markman (Eds.), Handbook of child psychology (Vol 3). NY: Wiley
- Gardner, H. (1983). Frames of mind: the theory of multiple intelligence. NY:Basic Books.
- Glaser, R. (1984). Education and thinking: the role of knowledge. American Psychologist, 39, 93-104.

Philippine Journal of Educational Measurement • Volume 10

Glaser R. (1986). Intelligence as acquired proficiency. In R. J. Sternberg & D. K. Detterman (Eds.), What is intelligence? Contemporary viewpoints on its nature and development. Norwood, NJ: Ablex Publishing.

- Glaser, R. (1991). Expertise and assessment. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition*. Englewood Cliffs, NJ: Prentice Hall.
- Herman., J. L. (1991). Research in cognition: Implication for achievement testing practice.
- Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. Cambridge: Cambridge University Press.
- Sternberg, R. J. (1988a). Mental self-government: a theory of intellectual styles and their development. *Human Development*, 31, 197-224.
- Sternberg, R. J. (1988b). The triarchic mind: a new theory of human intelligence. NY Viking Penguin Inc.
- Weinstein, C. E. & Meyer, D.K. (1991). Implications of cognitive psychology for testing: Contributions from work in learning strategies. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition*. Englewood Cliffs, NJ: Prentice Hall.

## TECHNIQUES OF ASSESSING, TYPES, AND TAXONOMY OF MEASURES OF THINKING SKILLS

Richard dlc. Gonzales De La Salle University Philippines

"Thinking is an art, with its own purposes, standards, principles, rules, strategies. And it is an art well worth learning, for every important thing we do is affected by our habits of mind".

Vincent R. Ruggiero, 1984

Teachers always have twofold tasks to accomplish once instruction begins. First, teachers must make sure that they deliver the instructional sequence that they have planned. Second, they must steadily assess the progress and success of their instruction in order to modify if necessary (Doyle, 1986). Monitoring of instructional progress is focused on determining whether planned activities are implemented while ascertaining success of instruction is done through classroom testing. Hence, one of the main concerns in the teaching-learning process is the quality of assessment tools used by classroom teachers.

Although researchers have provided numerous explanation of classroom achievement of students (Striggins & Bridgeford, 1985; Natriello, 1987; Gullickson &, Hopkins, 1987), it cannot be dened that the assessment tools contribute to the understanding of students' achievement and thinking skills. The assessment of student achievement and thinking skills is a factor in an instructional program that can not be neglected because of its importance in the teaching-learning process.

Tyler (1949) and more recently Popham (1990-1991) underscore that the main purpose of education is to change student behavior in a planned way. Both authors stress that instructional process is composed of three general steps. The first step encompasses planning of instruction, that includes identifying desired student behavior changes, selecting materials, and organizing learning experiences into a coherent, reinforcing sequence. The second step involves the instruction portion. Finally, the third step embodies determining whether the planned content of instruction has been covered and the student thinking skills have been developed. The last step makes sure that the students have learned the topics to be included and skills that need to be acquired (Airasian, 1994). Hence, this makes the last step the main basis for evaluating the worth of an instructional program which makes it imperative that classroom assessment tools should be of the highest quality besides being reliable and valid measures.

Philippine Journal of Educational Measurement • Volume 10

Previous studies clearly show that assessing students' thinking skills after an instructional program takes on several dimensions. As such, Stiggins, Griswold, and Wikelund (1989) attempted to summarize relevant studies on classroom assessment. They made references to the studies done by Stiggins and Bridgeford (1985) on quality control problems with classroom assessment based on teacher observation and judgment; Fleming and Chambers (1983) on problems with the quality of teacherdeveloped paper-and pencil tests, Schafer and Lissitz (1987); and Gullickson and Hopkins (1987) on inadequacy of the measurement training offered by educators.

In addition to these studies reviewed by Stiggins et al., it is noteworthy to mention the study conducted by Reynolds and Menard (1980) on test development practices of nearly 200 high school teachers along Bloom's taxonomy of educational objectives. Likewise, the study made by Fleming and Chambers (1983) which focused on teacher-developed paper-and-pencil test must not be forgotten.

In the Philippines, notable studies on measuring thinking skills were done in the natural sciences using Bloom's taxonomy of educational objectives as framework. The most recent one was done be España (1996) on the cognitive levels of the multiple-choice mastery test items in selected SEDP III textbooks. España found out that most items do not measure higher order thinking skills as they were limited to measuring memory skills. The findings of this study did not differ from earlier studies done by Deroga (1980) Sustento (1973) and Guco (1975).

Other studies also found out that tests cited in local textbooks measured only lower levels of thinking skills (Galvante, 1985; Gallos, 1987; Bentillo, 1990).

In sum, most teacher-made tests in the elementary and secondary levels, as can be found in foreign as well as local studies, measure only rote memory skills and a sprinkling of comprehension and application skills.

This paper, therefore aims to examine the measures of thinking skills in tertiary level, particularly in the Social Sciences employed by selected teachers of a provincial private tertiary institute in the Philippines.

This study identifies the techniques used in assessing thinking skills, the type of items or questions asked, and the level of cognitive taxonomy measured by each item or question.

### Taxonomy of Educational Objectives and Framework of Thinking Skills

The most commonly assessed behavioral domain in schools is the cognitive domain. This behavioral domain includes a range of thinking skills such as memorizing, interpreting, applying, problem-solving, reasoning, analyzing, and thinking critically. These behaviors that fall under cognitive domain have been organized into a taxonomy or a system of classification.

Bloom's Taxonomy of Education Ojectives. Among the taxonomies of educational objectives, the Bloom's Taxonomy of Educational Objectives: Cognitive Domain (Bloom, Englehart, Furst, Hill, and Krathwohl, 1956) is the most commonly used and referred taxonomy. Bloom et al uphold that cognitive domain is organized into six levels, with each level representing a more complex type of cognitive thinking skill or behavior. The level of cognitive domain of Bloom starts with the simplest and progresses to the most complex. The six levels are as follows:

- 1. *Knowledge*: Remembering previously learned material or memorization behaviors such as memorizing mathematical formulae, literary pieces, spelling of words, or events in the world history.
- 2. Comprehension : Grasping the meaning of material or understanding behaviors such as digesting or summarizing what one has read or explaining an idea in one's own words.
- 3. *Application* : Using information to solve unfamiliar problems and in concrete situations.
- 4. Analysis : Breaking down a large body of information into smaller parts.
- 5. Synthesis : Putting bits of information together into a whole or into a conclusion.
- 6. *Evaluation* : Judging the worth of a phenomenon, object, or idea for a given purpose using definite criteria.

Most of the studies on thinking skills have used this taxonomy in analyzing patterns of students' achievements and levels of cognitive behavior (España, 1996; Gallos, 1987: Bentillo, 1990).

Quelmallz's Taxonomy of Cognitive Behaviors. A more recent cognitive taxonomy that is used in many classrooms is proposed by Quellmalz, (1985). Quellmalz's taxonomy of cognitive behaviors is similar to Blooms taxonomy but

Philippine Journal of Educational Measurement • Volume 10

only includes five categories : recall, analysis, comparison, inference, and evaluation.

Although taxonomies differ in the particular levels or categories they include, Airasian (1994) argues that they generally distinguish behaviors that measure lower-level and higher-level cognitive domain. The two taxonomies reviewed stress that rote memorization, recall, and comprehension are skills that measure lower order thinking skills (LOTS). The behavior of comparing, inferring, analyzing, synthesizing, and evaluating belong to the higher order of thinking behavior or skills (HOTS).

This study employs the framework of thinking skills proposed by Quellmalz (1985). This framework was chosen because it combines the perspective of both philosophers and psychologists as they relate to critical thinking, problem solving, and cognitive processes. Table 1 on the following page shows the summary of Quellmalz Taxonomy of Educational Objectives.

#### Methodology

This study employed the descriptive type of research. It was directed towards determining the techniques of assessment, types, and taxonomy of measures of thinking skills in the social sciences at the tertiary level.

Data collection included the analysis of sample test papers, interview, and unstructured group discussion.

Analysis of sample test papers. The primary data were gathered from the sample of final tests from selected 15 teachers of Social Sciences of a provincial tertiary institution during the school year 1996-1997. These 15 teachers were teaching courses such as General Psychology, Introduction to Sociology, Philippine History, Philippine Government, Basic Economics and Taxation and Agrarian Reform. A total of 1,215 items was analyzed across the 15 respondent teachers.

The sample test papers consisted of mid-term or periodic examinations and long quizzes. There were 26 test samples included because only those that included more than 20 items were considered in the analysis. Each test consisted an average of 50 items. Items that were judged to be ambiguous, that is, unclear about what they measured were removed from the pool of items. The taxonomy proposed by Quellmalz was used to review and classify the test ites under investigation.

#### Bloom's Illustration Classification Definition Equivalent In this level, the students are required to recognize or remember Who invented Knowledge key facts, definition, concepts, rules, and principles. Recall the sewina and They attain these skills through rehearsal and practice. machine? Comprehension In this operation, students divide a whole into component/ whole relationships and the parts of cause/effect relationships What that characterize knowledge within subjects domain is are the different essential components of more complex tasks. The Analysis Analysis components can be the distinctive characteristics of features of objects or ideas, or the basic actions or procedures or the story? events. This level of thinking skill requires students to identify or explain How does similarities and differences. Simple comparison requires this A Comparison attention to one or a few very obvious component processes, element Synthesis while complex comparison requires identification of the differ from differentiation among many attributes or component actions. **B** element In this level, both deductive and inductive reasoning is employed. In deductive reasoning, the students are given a generalization What and are required to identify or explain the evidence that relates might be Application to it. In inductive reasoning, the students are given the evidence the good and Inference or details and are required to come up with the generalization. title for Synthesis Hypothesizing, predicting, concluding, and synthesizing all this story? require students to relate and integrate information. This level requires the students to judge quality, credibility, Is this a good worth, or practicality. Generally, students are expected to material for Evaluation use establish criteria and explain how these criteria are or a furniture? Evaluation are not met. The criteria might be established rules of Why or evidence, logic, or shared values. why not?

#### Table 1. Summary of Quellmalz Taxonomy\*

\*(Partly taken from Stiggins, Griswold & Wikelund, 1989).

*Interview and unstructured group discussion*. The other dataset used in this study was gathered through informal interviews and group discussion during a workshop on classroom test development conducted by the researcher.

The interview question was focused on how the teachers attempted to measure students' thinking skills in the courses they taught during the past two semesters. The discussion was done informally during the seminar conducted by the author.

#### Results

#### Techniques of Assessing Thinking Skills

There are three generally accepted techniques of assessing thinking skills and other classroom achievement. These are: 1) paper-and-pencil technique, 2) observation, and 3) oral-communicative technique.

The paper-and-pencil technique refers to assessment methods in which the students are asked to write down their responses to questions or problems. This is commonly known as test. The observation technique is an approach to gather data on how students think or behave cognitively which involves watching students perform activities and listening to students during classroom interactions and discussions. The last technique, oral-communicative is an approach where students are asked to answer questions and problems orally or verbally.

The data gathered from this study clearly show that paper-and-pencil technique or test is still the commonly used technique of assessing students' thinking skills and achievement. Although the respondent teachers mentioned that oral-communicative technique is often employed, it is not perceived as systematic and reliable compared to paper-and-pencil technique. The respondents disclosed that long tests as well as short tests or quizzes are given periodically to determine students' achievement and thinking skills. They added that students are more comfortable with this kind of assessment technique than the oral-communicative technique besides, the oral-communicative technique requires more time.

Observation as a technique in assessing thinking skills was not mentioned as being commonly employed. This is probably because, tertiary education does not allow so much time to conduct close observation of students' behavior because students at this level are supposedly behaved and more independent.

#### Types of Measures of Thinking Skills

Classroom assessment is almost synonymous to testing. In fact, there is a myth that assessment, measurement, and testing mean the same thing.

In terms of types of measures of thinking skills, the data were limited to the sample tests reviewed for this study. Types of measures refer to item types or format of which the questions or problems are asked and presented to the students.

Table 2 presents the results of the examination of the samples of tests included. It shows that the teachers in the study used a combination of selected-

response (59%), constructed-response (37%), and problem items (4%).

Туре	Frequ	uency	Perce	entage
Selected Response		717	and the second	59%
True-False	377		31%	
Multiple-choice	316		26%	
Matching	24		2%	
Constructed-Response	Tanka Heres	449	Personal State	37%
Completion	413		34%	
Essay	36		3%	
Problem Items	1. CP-1	49		4%
Case Items	12		1%	
Problem Solving	37		3%	
TOTAL	12	215	10	0%

Table 2. Distribution of questions/problems according to item types

The types of items that were asked and included in the tests are mostly of the completion (34%), true or false (31%), and multiple-choice (26%) item formats. Notably, however, in all the test samples, an item or two of the essay type were included (3%), usually placed at the last part of the test. Few problem solving items (4%) were included probably because of the nature of the subject matter. Surprisingly, the matching item type was not commonly used by the teachers in assessing thinking skills.

#### Taxonomy of Measures of Thinking Skills

Using the Quellmalz (1985) taxonomy of 5 classifications, Table 3 ppresents the distribution of items. The classification includes recall, analysis, comparison, inference, and evaluation.

Table 3. Distribution of items according to taxonomy

Taxonomy	Frequency	Percentage
Recall	632	52%
Analysis	230	19%
Comparison	219	18%
Inference	85	7%
Evaluation	49	4%
TOTAL	1215	100%

It can be gleaned from the table that more than one half (52%) of the test items in the sample tests measure recall. Analysis items comprise 19% and comparison items constitute 18% of the item pool. Items measuring inferential and evaluative thinking were hardly included with only 7% and 4%, respectively. The results definitely show that lower order thinking skills are measured more than the higher order thinking skills.

In the case of oral examinations, it was noted that items measuring comparative and inferential thinking were more conveniently used than recall, analysis, and evaluative thinking skills. In addition to this, the respondents pointed out that they felt the importance of including a couple of essay question because of their belief that an essay question measures higher order thinking skills.

### Discussion

Assessing students' performance and achievement is an important component of the teaching-learning process. The manner, however, in which this is done poses a lot of issues and concerns. The finding that paper-and-pencil approach is still heavily employed than the other approaches of assessing thinking skills is not surprising. This is so, because measurement and assessment is always perceived to be completed through testing alone. School's evaluation programs rely heavily on tests. Even national and local testing boards in many countries resort to paper-and-pencil tests to determine students' achievement. Whether a test is scientifically constructed or not, it is accepted that it can offer a better option and a more reliable estimate of any trait or skill.

The finding that tests are more widely used than observation and oralcommunicative approaches implies a lot in terms of training teachers in the assessment of classroom thinking skills. Tests, if not properly constructed and developed may provide erroneous information that could lead to an inappropriate decision. Hence, training of teachers to develop tests for classroom use should be considered an integral part of school's in-service training programs. Moreover, teachers should not only be equipped with skills on test construction but they should also be given and exposed to other alternatives of assessing students' thinking skills.

Although paper-and-pencil tests are more favored assessment tools, it was noted that the distribution of types of items in the tests reviewed was uneven. It is however interesting to note that teachers have attempted to use a variety of techniques in assessing the thinking skills of students but most of the items that teachers included in their tests were select-response types such as truefalse, multiple-choice and matching items. This means that teachers go through preparing so many items that may not necessarily measure higher-order thinking skills. True-false and multiple-choice items, for instance, may permit and encourage guessing and restrict processing of information to given options. Also, since alternative resposes are provided, a student may devote more time in reading them, if not guessing them, than to think of possible responses.

The data also show that constructed-response items only compose more than one-third of the items from relatively easy completion or short-answer items to difficult essay. Notably, almost all the test samples reviewed included at least one essay item. This was so, because teachers still adhere to the myth that a test is not complete without an essay item. Sad to say, however, that the essay items included did not go beyond measuring knowledge as most items required definition of terms and describing some facts.

Problem items were noticed, but they did not contribute significantly to the assessment of thinking skills because they only comprised 4 percent of the test items reviewed.

On the taxonomy of measures of thinking, the lower-order skills is evidently more frequently tapped than the higher ones. This is supported by the fact that more than half of the items only measured recall of fact, information, concepts, definitions, and rules. This result confirms earlier studies made locally and abroad (Stiggin, et al., 1989; Espana, 1996) that the higher order thinking skills are rarely and sparingly assessed even in the tertiary level.

The higher order thinking skills of inferring and evaluating were almost not observed particularly in the true-false items. The 49 items that measure evaluative skills were noted in some of the essay and problem-solving questions.

From the observations derived from this study, it can not be denied that students' thinking skills are not well tapped considering the fact that the type of items to measure them are relatively easy and the level of thinking is limited to the lower order. Such findings may create some anxiety to the educators today as seemingly schools are only developing lower order thinking skills of students. If the thinking process is limited to knowing of facts, undeniably, students will never develop the skills to compare, infer, analyze, and evaluate. The goal therefore of schools knowing this fact is to develop curricula and testing programs that will enhance the higher order thinking skills of students.

Knowing that this study was done in a higher institution of learning, one would expect that tests differ from those in the basic education which is only focused on knowing the ability of its students. In addition to this, considering the nature of the courses taught where these tests were given, one would likely believe that more demanding questions should have been included. Social

Philippine Journal of Educational Measurement • Volume 10

sciences such as psychology, sociology, economics, and government are subjects that should involve more time for analysis, comparison, and evaluation. It was clearly indicated that teachers did not have attempt so much to measure higher order skills as they only focused what students were able to memorize.

Another factor why such data were yielded was the fact that teachers who taught these subjects may not have the necessary background in assessment in general and testing in particular. In the interview conducted, teachers expressed their desire to learn more on how to assess classroom thinking and showed so much interest in developing such skill. It appears then, that the function of assessment tool, the types, and the taxonomy is related to the teachers' educational background and training. Hence, this study clearly suggests that training teachers to assess classroom thinking skills is very much needed.

#### Conclusion

From the results, it was obvious that teacher-made-tests are still heavily employed by teachers to measure thinking skills. Creativeness of teachers in test development was not displayed because identification and true-or- false item formats are still prevalently used. The more effective item formats which could be multiple choice and others were not favored maybe because of the difficulty of constructing items or because teachers have limited knowledge and skills in developing such types of test format. The basic thinking skill - recall, which is purely rote memory, is still the dominant measure of thinking skill. This implies that teachers are not maximally tapping the thinking skills of students because they fail to include items that can tap and assess the higher thinking levels of students.

Although it was not the focus of this study to look into the mechanics of item writing, it was observed from the sample test items that a lot of rules on item writing and test development were violated.

In general, it can be said that testing is not necessarily the most effective tool in measuring thinking skills especially on tertiary level of education.

#### Recommendation

This study has a lot ofimplications on training of teachers in developing tests appropriately and creatively. More importantly, it is obvious that teachers need to further their understanding of assessing knowledge and thinking skills as well as the psychological. social and educational basis of assessing students. Training of teachers to measure thinking skills should be institutionalized especially on the tertiary level where most teachers have not undergone teacher training education. It is further recommended that measuring thinking skills should be given equal importance in planning and implementing instructional programs.

### References

Airasia, P.W. (1994) Classroom assessment. New York: McGraw-Hill, Inc.

- Bentillo, E. N. (1990). Summative evaluation of the textbook physics in your environment. *Monograph No. 43*, Quezon City : UP Institute of Science and Mathematics Education.
- Bloom, B.S., Englehart, M.D., Furst, E.J., Hill. W. H., and Krathwohl, D. R. (1956). Taxonomy of educational objectives: Handbook 1 Cognitive Domain. New York: McKay Publishing.
- Deroga, G. (1980). Levels and types of questions asked by teachers. Unpublished masters thesis, University of the Philippines, Quezon City.
- Doyle, W. (1986). Classroom organizational and management, In M.C. Wittrock (ed.). *Handbook of research on teaching*. (pp. 392-431). New York : MacMillan.
- Ennis, R. (1987). A taxonomy of critical thinking dispositions and abilities. In J. Baron and R. Sternberg (Eds.), *Teaching thinking skills*. NY: Freeman.
- España, R. C. N. (1996). The cognitive levels of the multiple-choice mastery test items in selected SEDP III textbooks based on Bloom's taxonomy. Unpublished doctoral dissertation, De La Salle University, Manila.
- Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows to the classroom. In W.E. Hathaway (Ed.) *Testing in the schools* (pp. 29-47). San Francisco, CA: Jossey-Bass.
- Gallos, M. R. (1987). The cognitive emphases of teacher-made chemistry tests in colleges of Metro Cebu. Unpublished masters thesis, University of thePhilippines, Quezon City.
- Galvante, W.P. (1985). Analysis of HSS 1 textbook exploring our environment for level of cognitive demand. *Monograph No. 36*. Quezon City : UP Institute of Science and Mathematics Education Development.
- Guco, S.W. (1975). Analysis of questioning techniques of selected elementary mathematics teacher and the implications of findings. Unpublished masters thesis, University of the Philippines, Quezon City.
- Gullickson, A. R. & Hopkins, K. D. (1987). Perspective on educational measurement instruction for preservice teachers. *Educational Measurement Issues and Practice*, 6(3), 12-16.
- Hopkins, C. D., & Antes, R. L. (1985). Classroom measurement and evaluation, 2<sup>nd</sup> Ed., Itasca, IL: F.E. Peacock Publishers, Inc.

Philippine Journal of Educational Measurement • Volume 10

Natriello, G. (1987). The impact of evaluation process on students. *Educational Psychologist, 22, 155-175.* 

Popham, W.J. (1990). Modern educational measurement. Englewood Cliffs. NJ: Prentice -Hall.

Popham, W.J. (1991). Appropriateness of teachers' test preparation and practices. *Educational Measurement: Issues and Practice*, 10(4), 12-15.

Quellmalz, E.S. (1985). Developing reasoning skills. In J. R. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice*. (pp 86-105). New York: Freeman.

Reynolds, W. M. & Menard, K. A. (1980). An investigation of teachers' test construction practices. Paper presented at the annual meeting of the National Council of Measurement in Education, Boston.

Ruggiero, V. R. (1984). The art of thinking: A guide to critical and creative thought. Cambridge: Harper & Row, Publishers.

Salvia, J., & Ysseldyke, J. E. (1991). Assessment. 5<sup>th</sup> Edition, Princeon, NJ: Houghton Mifflin Co.

Sermonia, T.C. (1982). An analysis of teacher-made tests. Unpublished masters thesis, De La Salle University, Manila.

Schafer, W.D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education*, 38 (3), 57-63.

Sternberg, R. (1988). Critical Thinking. In F. Link (Ed). Essays on the Intellect. Alexandria, VA: ASCD.

Stiggins, R. J., Griswold, M. M., and Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, Vol 26, No. 3, 233-246

Stiggins, R.J. (1988). The impact of teacher training on classroom assessment. Portland, OR : Northwest Regional Educational Laboratory.

Stiggins, R. J., Conklin, N.F., & Bridgeford, N.J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice*, 5(2), 5-17.

Sustento, S. A. (1972). Questioning practices of elementary science teachers. Unpublished masters thesis, University of the Philippines, Quezon City.

Tyler, R. W. (1949). Basic principles of curriculum and instruction. Chicago: University of Chicago Press.

## THE LEARNING ENVIRONMENT AND SUPERVISORY PRACTICES OF A PRIVATE SECONDARY SCHOOL

#### Zulkifli A. Manaf and Rahimah Haji Ahmad University of Malaya Malaysia

This paper presents the findings of a research done to assess the learning environment and the supervisory practices that promote instruction in a private secondary school in Kuala Lumpur. The assessment was carried out upon the request of the administrative personnel of the school. The purpose of the assessment is to identify the strengths and weaknesses of the school's learning environment with an intention to improve the situation. The main body of the paper deals with the perceptions of the teachers of the school learning environment as measured by the School Climate Profile instrument in terms of interpersonal relationship, teaching and learning, administrative leadership, and physical facilities. It also includes the teachers' perception of the supervision practices afforded by the leadership personnel of the school in terms of professional support, instructional support and staff development program. In addition to that, a list of suggestions from the teachers of the ways in which the leadership personnel of the school can help to maintain and promote instruction in the school is also presented. A discussion of the highlights and deficiecies of the school in terms of the variables of school climate and supervisory practices follows. A list of suggestions and recommendations for the school leadership personnel to improve the school is also presented.

This paper presents the findings of the assessment of the learning environment (school climate) and supervisory practices of a private secondary school in Kuala Lumpur, Malaysia. School effectiveness and school improvement have been widely researched all over the world over the past two decades (Brookover et al., 1979; Strivens, 1985; Mortimore et. al., 1988; Rutter et. al., 1979; Edmonds, 1979; Lezotte and Jacoby, 1991; Townsend, 1991; Reuter, 1992; Reynolds and Cuttance, 1992; McGaw and Others, 1993; Cheng, 1993; Lee et. al., 1993; Creemers, 1993; Franklin and Others, 1993; Reynolds et al., 1994; Hargreaves, 1995). In some of the effective school studies, school climate has also been studied and used as one of the measures of school effectiveness. The underlying notion is that we are treating the school as an organization and as such has its own personality, just like individuals. Hoy et al. (1991) defined school climate as follows:

Philippine Journal of Educational Measurement • Volume 10

"School climate is the relatively enduring quality of the school environment that is experienced by the participants, affects their behavior, and is based on their collective perception of behavior in schools." (p. 10)

What does that mean? Many researchers believe that this notion of school climate can help link characteristics of organizations with individual attitudes and behaviors and show how organizations affect individuals' behavior independent of their own characteristics. In other words the same individuals may behave differently in groups with different organizational climate.

How is organizational climate measured? It is done by assessing group member's perceptions of organizational characteristics. It is usually accomplished using standardized measures/scales to measure individual perceptions of their environment. Pioneers of this kind of measures of school climate are Halpin and Croft (1962) and Hoy et. al. (1991). According to Halpin and Croft (1962) the school climate is the feeling one gets when one is in the school compound.

#### Rationale of Study

The use of school climate as an indicator of school effectiveness is closely related to the internal process model postulated by Cheng (1996). According to Cheng (1996) one of the ways in which we can look at school effectiveness is to consider the internal processes that occur within the school. The internal process model assumes that a school is effective if its internal functioning is smooth and "healthy". It argues that to view effectiveness in terms of a process instead of an end state is a solution that at least minimizes many of the obstacles to effectiveness. Therefore, internal activities or practices in a school are taken as important criteria of school effectiveness. For example, the school climate indicators such as interpersonal relationships, teaching and learning and administrative leadership (Ahmad & Manaf, 1997a) fits very well in the internal process model. Leadership, communication, participation, co-ordination, adaptability, planning, decision making, social interactions, school climate, teaching methods, and classroom management are often used as indicators of effectiveness. Furthermore, according to Walberg and Greenberg (1997) in optimal environments, children and adults will enjoy themselves more and get more things done. They added that research has shown that the classroom social environment is one of the chief psychological determinants of academic learning.

#### Purpose of study

A visit was made to a certain private secondary school (identified only as School A- in the text from now on) on October 18, 1996 to study the school climate and supervision practices that promote instruction. The purpose of this study was to assess the overall situation as perceived by the teachers with long-term intention o improve it.

#### School Climate

To assess the school climate at School A, the School Climate Profile instrument was used. This instrument was adapted from an earlier version that was used in another study (Ahmad, 1981; Ahmad & Manaf, 1997a; Manaf, Ahmad, and Kim, 1995). The School Climate Profile instrument had four subscales: (a) interpersonal relationship (b) teaching and learning (c) administrative leadership and (d) physical facilities. Why are these components important in a school organization? The school climate is an important feature of any school because it tells us a lot about the interactions of the members of the school organization and its effect on student achievement. The school climate is somewhat related to the administrative leadership of the school. This in turn affects the teachers and student morale. Since a school is usually linked to teaching and learning, the school climate and the kind of supervision practiced in the school will affect the commitment of the teachers in their teaching. This in turn will tend to influence student achievement.

To be effective in promoting good teaching, the supervisory practices and leadership of the school must be in tandem with the total staff aspiration and needs. According to Slater and Teddlie (1992) in order for a school to be effective the following factors need to be present: (a) appropriate administration (b) teachers are prepared to teach, and (c) students are ready to learn. Appropriate administration is related to the provision of a conducive work place including direct support to teachers in their instruction such as guidance and encouragement. It is not only regarding physical conditions but includes also interpersonal relationships and co\_operation among teachers and students alike. The principal leadership skills play an important role in determining school effectiveness (Evans & Teddlie, 1993; Ahmad & Manaf, 1997b). Teacher preparedness does not imply merely on the teaching preparations done by the teachers to deliver their lessons but to a certain extent, the degree and level of commitment teachers are willing to put into their job. Usually, if teachers are prepared to teach then students are motivated and ready to learn. Undoubtedly this in turn will affect their academic achievements. If we have these conditions then we have the right ingredients in making any school become a world class school.

#### School climate and principal leadership

The importance of school climate and principal leadership in determining school effectiveness can never be over emphasized. According to Stockard and Mayberry (1992) the overall school climate or culture of a school which is influenced by effective leadership affects learning. Schools and classrooms that

Philippine Journal of Educational Measurement • Volume 10

enhance achievement appear to be characterized by high academic expectations, effective leadership, an orderly atmosphere and warmth, concern and respect for others. In other words, effective schools usually have good interpersonal relationships among the members of the organization. This feeling could be sensed when one is in the school compound especially in the teacher's lounge. The warmth and orderly atmosphere could also be measured using the perceptions of the school members through administration of an appropriate school climate instrument (Halpin and Croft, 1962).

Richards (1991) put forward the idea that schools did make a difference and they did so for reasons that could be traced directly back to alterable organizational characteristics of the school: a focus on basic education, directive leadership, high academic expectations, orderliness and a positive school climate. All these variables will distinguish effective schools from ineffective schools.

Administrative leadership is somewhat related to supervision practices that promote instruction in any school. In schools where the principal is regarded as an instructional leader, teaching and learning is strongly perceived by the teachers and students of the school in a positive way. (Ahmad & Manaf, 1997a; Garibaldi, 1993).

#### Supervision and supervisory practices

According to Glickman (1990) supervision can be thought of as a "glue" of a successful school. In that sense, supervision is sometimes referred to as the function in schools that draws together discrete elements of instructional effectiveness into whole-school action. Regardless of the schools diverse characteristics, successful schools have a common glue that keeps the staff together and creates consistency among the school's various elements. The glue referred to here is effective supervision. However effective supervision does not happen by chance. It requires knowledge, interpersonal skills, and technical skills. These in turn are applied through supervisory tasks of direct assistance to teachers, curriculum development, and staff development. This adhesive pulls together organizational goals and teacher needs; and provides for improved learning. Glue is used here as a metaphor. Incidentally, glue is a good metaphor for effective, fully functioning school supervision. However, in reality, successful supervision is usually taken for granted but when problem arises, blame is put in principal leadership. An ideal situation should be as follows: Teachers are in the forefront of successful instruction; supervision is in the background providing support, knowledge and skills that enable teachers to succeed. Therefore, when improved instruction and school success do not materialize, inept

supervision should be blamed for not permitting teachers to be successful. Who are the supervisors in the school? Any one responsible for improving classroom and school instruction is referred to as a supervisor.

The supervisor in this context refers to any person involved with supervision, and not to a particular title or position. However in most situation, the principal is usually the supervisor. Sometimes in certain schools or school system the supervisory function is empowered to other staff member such as the assistant principals, department heads, etc. In this sense, supervision is viewed as a process and function, and not a particular position or person. A common characteristic of successful schools is that someone, somewhere is responsible for and committed to the process, function, and tasks of supervision. To summarize, we can say that behind every successful school is an effective supervision program. Supervision practices relate, to a certain extent, to the leadership personnel of the school and what they do to support and help teachers in their work i.e., teaching. To facilitate such collective instructional improvement, those responsible for supervision must have certain prerequisites. These are (a) knowledge base; (b) interpersonal skills base, and (c) technical skills. According to Glickman (1990) supervisors have certain educational tasks that enable teachers to evaluate and modify their instruction. In planning each task the supervisor needs to plan specific ways of giving teachers a greater sense of professional power to teach students successfully. Among those supervisory tasks that have such potential to affect teacher development are: (a) direct assistance; (b) curriculum development, and (c) staff development. These are similar constructs that were measured by Part III of the instrument used in the study.

The instrument measured the extent of the leadership personnel's contribution in promoting instruction in the school. Supervision, in this context was not confined to a narrow scope of overseeing teachers' activities with the intention of evaluating their performance. It also include other aspects such as whether the leadership personnel provide teacher support to increase their teaching proficiency (professionalism), help teachers in classroom instruction and to what extent the staff development program provides avenues for professional development.

#### Research Questions

,

r

E

1

t

Э

Э

,

Э

1

t

3

3

;

,

ł

f

2

ţ

- 1. What is the school climate like at School A?
- 2. What are the characteristics of the supervision practices at School A?

#### Methodology

#### Sample

All the teachers in the school were used as subjects of this study. A total of 79 teachers from all levels i.e., the kindergarten, primary and secondary, participated in this study. The instruments were administered after a short presentation was given in the school hall. The researchers explained the importance and significance of this research and assured anonymity of their responses. It was also stressed that the questionnaire was not a test of any sort and that each person's responses should be independent of one another in the group.

#### Instrumentation

The instrument has three parts. Part I consists of questions regarding the teacher's background information such as sex, race, age, academic qualification, professional qualifications, total teaching experience, teaching experience in the school, major area of specialization while in college or university, and level of satisfaction regarding the teaching profession, teaching in present school, and teacher preparation program.

Part II contains 30 items which measure the school climate. It has 4 subscales, namely; (a) interpersonal relationship (b) teaching and learning (c) administrative leadership and (d) physical facilities.

It consists of items describing the situations and processes that take place in the school. The teachers are instructed to respond to each of the items on a fourpoint scale. The value of the scale categories are as follows: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.

The instrument was adapted and translated from an English version (Ahmad & Manaf, 1997a). It has face validity and content validity based on the content analysis of the items used. The internal consistencies, as measured by the standardized Cronbach alpha coefficient, vary from 0.77 to 0.90. For all practical and research purposes these values are desirable and acceptable.

Part III of the instrument contains items that measure supervision practices that promote instruction. It has three sub-scales, namely; (a) professional support, (b) instructional support, and (c) staff development program. The professional support subscale contains 11 items which are related to the extent of professional support to teachers afforded by the leadership personnel of the school.

The instructional support sub-scale contains another 17 items that cover the extent of support given by the leadership personnel to help teachers in their classroom instruction.

Finally, the staff development program sub-scale contains another 16 items that measure the extent to which the program has helped them in their profession. This instrument was adapted and translated into the Malay language by the researchers from an English version (Jones, 1986). The authors produced two versions of the translations. The two versions were compared to the original English version and the items that best convey the meanings intended by the original author were used in the final instrument. The respondents in the study (teachers) indicated that they understood the meaning of the items when it was administered. This indicate content validity of the items. The internal consistencies of the items in the three sub-scales are presented follows:

Sub-scale	No. of Items	Cronbach Alpha
Professional Support	11 items	0.93
Instructional Support	17 items	0.94
Staff Development Program	16 items	0.94
Supervision (Total)	44 items	0.97

These values indicate a high degree of internal consistencies of the items in each subscale which consequently lend a high degree of reliability to the instrument.

#### Results

The results are presented in two parts. Part 1 deals with the school climate while Part 2 deals with the supervisory practices that promote instruction. However the discussion for this paper is mainly focused on those items where the percentage of teachers who disagree to the items presented is relatively high. Those items should merit the attention of the leadership personnel of the school. This is presented in Table 1.

#### School Climate

On the whole, the school climate is positive. However there are certain aspects of the school climate that need attention and improvement. The discussion is based on the related items presented in Table 1 where the percentage of respondents expressing disagreement to certain items

Philippine Journal of Educational Measurement • Volume 10

are quite high. Attention is drawn to the following climate items in the related sub-scales:

Table 1. School climate p	profile (N=79	)
---------------------------	---------------	---

ter al melle began est en anne el si el compañe de la c	Disagreem		
Description of the Item		%	
Teachers in this school are enthusiastc.	21	26.6	
I feel that some of my ideas are being practiced in the school.	15	19.2	
School decision are made after discussion with party concerned.	33	41.8	
I have an opportunity to voice up my concerns.	23	29.1	
School environment encourage students to be co-operative.	19	24.1	
In this school we always "Look Forward".	28	35.4	
I feel comfortable in the school.	27	34.2	
This school has adequate space for teaching and learning.	57	73.1	
This school has adequate facilities for satisfactory teaching and learning.	23	29.5	
Everyone is proud of the physical facilities of the school.	31	39.7	

Interpersonal relationship. Generally speaking, interpersonal relationships among some of the teachers was not very encouraging (21.5%). Further it is noted that 34 percent of the teachers felt that they are not proud to be teachers of this school. This is not healthy because if this situation is allowed to persist and increase, then the school climate and learning environment will be eroded over time. The school climate will affect the school effectiveness in some way or another. There is a lot of things that the leadership personnel could do to promote better understanding among the teachers and other staff members, including the leadership personnel, so that they can feel like being part of a big family and be proud members of the school family. Teachers should be inculcated with a team spirit and undivided loyalty towards the school. They should develop a sense of belonging to the school. A more open staff meeting on a regular basis where teachers can voice out their opinion and suggestions should be considered. A faculty club monthly activity and family day may improve the situation. This is one of the characteristics of effective and 'improving' government-run schools that have been studied elsewhere.

Administrative leadership. As far as the administrative leadership is concerned, the single outstanding characteristic of the administrative personnel is the lack of open door policy or two-way communication with the teachers. Even though the administration may not wish to involve the teachers in certain policy matters regarding the governing of the school, the teachers feel that they need to be consulted regarding matters that affect their well-being in the school working environment. There should be an open-door policy where teachers are free to convey their wishes and/or grievances to the administration. This practice is consistently found in the more effective schools studied (Ahmad & Manaf, 1997a; 1997b).

*Physical facilities.* Some of the teachers perceived that there are some deficiencies in the physical facilities in the school. About 34 percent of the teachers do not feel comfortable in the school because of its general state of maintenance. If that is true then the leadership personnel should look into this matter and institute corrective measures to improve the situation. Probably these teachers have a more stringent criteria for cleanliness. Again the spins of belonging to the school (or lack of it) is evident here. high percentage of teachers (39.7%) felt that they are not proud of the school facilities let alone wanting to keep it in good condition. This attitude among teachers need to be improved through better collaboration and understanding between administrators and teachers.

#### Supervision Practices That Promote Instruction

*Professional support*. The discussion regarding professional support is based on the items presented in Table 2.

		Rarely
Description of the Item	freq	%
Assistance to identify and procure materials for instruction	25	31.6
Demonstration of classroom teaching technique	43	55.1
Consultation regarding teaching problem	30	38.5
Opportunity to schedule and plan observation	27	34.2
Observation to improve instruction technique	26	33.3
Discussion after an observation session to discuss and analyze the lessons observed	37	47.4
An opportunity to analyze lesson with the help of audio and video tapes	41	54.7
Opportunity to observe and discuss a class taught by another teacher	53	68.8
Concern for teachers and students	30	38.5
Giving support and assistance where needed	21	26.6
Give encouragement to think about matters regarding teaching skills	19	24.4

Table 2. Supervisory practices that promote in struction (N=79)

Philippine Journal of Educational Measurement • Volume 10

As summarized in Table 2, it is evident that the teachers felt they need more professional support as indicated by the respective items in the sub-scale. According to the teachers, the professional support as measured by these items occur very rarely.

It is observed that the teachers felt that they need more professional support from the leadership personnel to improve their teaching competency. Probably this is due to the fact that about 50 percent of the teachers at School A did not have any formal teacher education program.

*Instructional support.* The teacher's perceptions regarding instructional support were presented in Table 3.

Description of the Item		Rarely
		%
Planning suitable teaching objectives	25	31.6
Designing suitable teaching activity	28	35.4
Planning remedial or enrichment activities	29	37.7
Plan activities to study and diagnose	38	48.1
Plan learning strategy to group learning	33	42.3
Plan additional learning activity for student who complete their work quickly	36	46.2
Plan and organize space and materials for instruction	24	30.8
Control student behavior (discipline)	28	35.9
Explain classroom rules and regulation to student	26	33.3
Gives clear direction and prepare student for transitions	22	28.2

Table 3. Instructional support - curriculum development (N=79)

With reference to Table 3, about half of the teachers felt that they need more instructional support in curriculum development as reflected by the items concerned so that they can teach better in their classroom.

Professional development program. The teachers' perceptions regarding professional development program were presented in Table 4.

Referring to Table 4, the teachers are in dire need of a more organized staff development program. While most of them felt that the staff

development program is very beneficial for their professional development there are two aspects that need further attention. These are managing activities through demonstration and the access to materials to implement the program at the end of the formal presentation.

		Rarely
Description of the Item	freq	%
Opportunities to plan together an in-service activity	27	35.1
Learning activities which give knowledge and skills that are proven to be effective	21	27.3
Presentation of knowledge and skills that are practical and useful	25	32.5
Well planned activities	29	37.7
Knowledgeable and credible facilitator	23	30.3
Opportunity to learn to solve problem and interact in small groups	27	35.5
Programs that extend for several sessions	28	36.8
Presentations that use a variety of teaching techniques	25	32.9
Management of activities through demonstration (live or recorded)	42	55.3
Opportunity to discuss and share ideas about knowldge or skill learned	25	33.8
Opportunities to use and apply knowledge learned	23	30.3
Opportunity to observe and get feedback in between sessions	26	34.2
Observation by leadership personnel to ensure that knowledge learned is put into practice	29	38.2
Facilities to get materials and resource persons at the end of the program	31	40.8

Table 4. Staff development program (N=79)

Basically, the teachers need more supervisory support so that they can be more professional and productive in their classroom instruction. In addition to that there should be more interaction among the teachers and two-way communication between the leadership personnel and the teachers. To summarize, it can be said that the school climate, school culture, and supervisory practices at School A were on the positive side but there is room for improvement. The improvement suggested was based

Philippine Journal of Educational Measurement • Volume 10

on the teachers' responses on the specific items in the school climate profile and supervision practices.

Suggestions by teachers to increase instruction. When asked to give suggestions on what ways can the leadership personnel (school manager) can help the teachers to improve instruction, the teachers provided a variety of responses. These are presented and organized in Table 5. Basically, the teachers need more say in decision making that occurs in the school especially those that affect their work. Further they were requesting for a better working condition, benefits, and better prospects in their career development.

Table 5. Teachers' suggestions to maintain and improve instruction

Suggestion	Frequency
In decision making, teacher's views and suggestions must be taken into consideration	5
A stricter disciplinary rule should be enforced	5
More information should be channeled to teachers regarding the latest development in education published in magazines, journals and newspapers	4
Be more receptive to teacher problems when dealing with problem students	8
Give incentives to teachers, e.g., 20-30% increase in salary, yearly bonus, housing allowance and loan	19
Being more professional by taking into considerations teachers interest in whatever decision making	8
Appreciate teachers contribution towards students	4
Provide teachers with audio visual aids	8
Counseling services for teachers	5
Sponsorship for teachers to further their education	6
Student entry should be based on a special admission test	3

#### Recommendations

#### School climate.

The leadership personnel should consider the leadership style by being more democratic. The principal should try to institute an open-door policy for staff

members. The teachers should be involved in major decision-making especially in those matters that have implications on them.

#### Professional support

The school leadership personnel should consider seriously organizing a workshop/seminar seminar in the form of in-service training to upgrade the teachers' professional knowledge and skills in general pedagogy and specific teaching techniques. As indicated by the teachers, the administration should also consider having counseling clinics for teachers.

#### Instructional support

The school leadership personnel should consider seriously organizing a workshop/ seminar in the form of in-service training to upgrade the teachers professional knowledge and skills in planning activities, identifying problems and making diagnosis, plan for group learning strategy and planning activities for fast learners.

#### Staff development

The school leadership personnel should consider a workshop for staff in managing activities through demonstration (live or recorded). The leadership personnel should ensure that workshop materials and resource persons should be readily accessible to teachers in order to implement any program presented in the workshop.

#### Conclusion

The school climate and supervision practices at School A is on the positive side but in order to improve the condition to take the school effectiveness to greater heights the school leadership personnel needs to take the initiative to carry out the recommendations that were put forward based on the findings of the preliminary research conducted in the school.

#### References

- Brookover, W., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). School social systems and student achievement: Schools make a difference. New York: Praeger.
- Cheng, Y. C. (1993). Profiles of Organizational culture and effective schools. School Effectiveness and School Improvement, v2, 85\_110.

Philippine Journal of Educational Measurement • Volume 10

Cheng, Y. C. (1996). *The pursuit of school effectiveness*. Hong Kong: Hong Kong Institute of Educational Research.

- Creemers, B. P. M. (1993). Towards a theory on educational effectiveness. Paper presented at the Annual Meeting of the International Congress for School Effectiveness and Improvement, Norrkoping, Sweden, January 1993.
- Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37, 15\_24.
- Evans, L. & Teddlie, C. (1993). Principals change facilitator styles in schools that differ in effectiveness and SES. Paper presented at the annual meeting of the American Educational Research Association at Atlanta, Georgia, USA., April 12\_16, 1993.
- Franklin, B. J. and Others. (1993). Measures of school effectiveness; Consistency / Inconsistency when models are varied across subject area tests. Paper presented at the Annual meeting of the American Educational Research Association, Atlanta, Georgia, U.S.A., April 12\_16, 1993.
- Garibaldi, A. M. (1993). Improving urban schools in inner city communities. Occasional paper #3. Urban Child Research Centre. Cleveland State University, Ohio, USA
- Glickman, C. D. Supervision of Instruction: A developmental approach (2nd Ed.). Boston: Allyn and Bacon.
- Halpin, A. W., & Croft, D. (1962). The organizational climate of schools. Washington DC: US Office of Education, Dept. of Health Education and Welfare.
- Hargreaves, D. H. (1995). School culture, school effectiveness and school improvement. School Effectiveness and School Improvement, 6(1), 23\_46
- Hoy, W. C., Tarter, C. J., & Kottkamp, R. B. (1991). Open Schools/Healthy Schools: Measuring organizational climate. Newbury Park, CA: Sage.
- Jones, J. W. (1986). A data collection for describing research\_based supervisory practices for promoting instructional improvement in a. local school district. Unpublished Ed. D dissertation, University of Georgia, USA
- Lee, V. E., Bryk, A. S. & Smith, J. B. (1993). The organization of effective secondary schools. Chapter 5 in Darling\_Hammond, L. (Ed). Review of Research in Education. Washington DC, American Educational Research Association pp. 175\_267.
- Lezotte, L. W., & Jacoby, B.C.(Eds.).(1991). *Effective Schools practices that work*. Effective School Products Ltd. ERIC Document Reproduction Number ED 359649.
- McGaw, B. and Others (1993). Making schools more effective: Report of the Australian Effective Schools Project. ERIC Document Reproduction Number ED 361829.
- Mortimore, P., Sammons, P., & Stoll, L. (1988). School matters. Somerset :Open Books.
- Rahimah Haji. Ahmad, & Zulkifli A. Manaf (1997a). The Characteristics of effective and less effective schools in rural and urban settings: A case study

from Malaysia. Paper presented at the 10th International Conference on School Effectiveness and Improvement, Memphis, Tennessee, USA, 5\_8 January, 1997.

- Rahimah Haji. Ahmad, & Zulkifli A. Manaf(1997b). Principal leadership and school effectiveness. Paper presented at the 10th International Conference on School Effectiveness and Improvement, Memphis, Tennessee, USA, 5\_8 anuary, 1997
- Reuter, S. F. (1992). Characteristics of successful schools: Perception differences between rural and urban elementary schools teachers. Paper presented at the 14th Annual Rural and Small Schools Conference at Manhattan, Kansas, U.S.A., October 26\_27, 1992. ERIC Document Reproduction Number ED 360115.
- Reynolds, D., Creemers, B. P. M., Nasselodt, P. S., Schaffer, E. C., Stringfield, S., & Teddlie, C. (1994). Advances in School Effectiveness Research and Practice. London, New York and Tokyo: Pergamon.
- Reynolds, D.,& Cuttance, P.(eds.)(1992). School Effectiveness. London and New York: Cassel.
- Richards, C. E. (1991). The meaning and measure of school effectiveness. In Rethinking effective schools research. James R. Bliss, William A. Firestone & Craig E. Richards (Eds.). Englewood Cliffs, NJ: Prentice\_Hall.
- Rutter, M., Maughan, B., Mortimore, P., Ousten, J., & Smith, A., (1979). Fifteen thousand hours: Secondary schools and their effects on children. Cambridge, MA: Harvard University Press.
- Slater, R. O. & Teddlie, C. (1992). A theory of school effectiveness and leadership. School Effectiveness and School Improvement, 3 (4), 242\_257.
- Strivens, J. (1985). School climate: A review of a problematic concept, in D. Reynolds (ed.), Studying School Effectiveness, London and Philadelphia: Falmer Press.
- Stockard, J., & Mayberry, M. (1992). Effective educational environments. Newbury Park, CA: Corwin Press Inc.
- Townsend, T. (1991). School effectiveness: A View from the school. Paper presented at the Annual Meeting of the 4th International Congress for School Effectiveness and Improvement, Cardiff, Wales, United Kingdom.
- Walberg, H.J., & Greenberg, R. C. (1997). Using the learning environment inventory. *Educational Leadership*, 54 (8), 45\_47.
- Zulkifli A. Manaf, Rahimah Haji Ahmad, & Lee Ong Kim (1995). An Evaluation of the school climate profile instrument. Paper presented at the 2nd Asia Pacific Conference on Educational Assessment and Research, Chiang Mai, Thailand, 4\_8th September 1995.

## FACTORS RELATED TO STUDENT RATINGS ON TEACHER EFFECTIVENESS

#### Simon Lloyd D. Restubog De La Salle University-College of St. Benilde Philippines

This study is an attempt to determine the degree of relationship of student ratings on teacher effectiveness with specific teacher, student, and course characteristics. Likewise, it sought to determine which among the specific characteristics are predictive of student ratings. The Student Instructional Report, an instrument designed to measure teacher effectiveness, was employed for data collection. This was administered to four hundred classes in a community college in Metro Manila. Product moment correlation and stepwise regression were used for data analysis. Results showed that teaching methodology, perceived strictness-leniency, students' expected grade, level of interest, and perceived level of course difficulty are significantly correlated with student ratings on teacher effectiveness. Further analysis showed that teaching methodology, perceived strictness-leniency, students' level of interest and perceived course difficulty emerged as significant predictors of student ratings. The implications of the results in terms of theoretical significance and integration with relevant literature were also discussed.

What is considered as effective teaching? Research findings show that the key to success in teaching must rest on three important teacher characteristics (Medley in Woolfolk, 1995). First is knowledge which refers to knowing the important facts, concepts, principles and aplications in an academic field or discipline. Teachers ought to know how their knowledge can effectively be simplified and integrated into the curriculum. Then, explained into bits and pieces of information, such as lessons, discussions, demonstrations, and all the other activities of instruction. Second is organization which refers to the actual presentation of the lesson which should be presented in such a way that students will clearly understand and apply the concepts that they have learned. The third is warmth and enthusiasm which according to some studies have found a positive relationship with students' achievement (Rosenshine & Furst, 1973). This implies that teachers who are warm and friendly tend to have students and classes who like them, in general.

While many empirical studies seem to concur that knowledge of subject matter, having a well-planned and organized lecture, and ability to inspire confidence and interest among the students are the three most important characteristics of effective teaching, some reports accent that teacher effective-

ness is closely related with certain teacher personality traits (Meredith, 1988; Crawford & Bradshaw, 1968; Gadzella, 1968; Musella & Rusch, 1968; French, 1957; and Coffman, 1954). Such traits as cooperativeness, kindness, respect for the learner, good sense of humor, fairness, enthusiasm, and an accepting attitude towards student views appeared as common qualities of good teachers (Steffens, 1990; Endle, Murray, & Rushton, 1985; Waters, Kemp, & Pucci, 1988; Hamachek in Travers & Rebore, 1995; Darshan, 1976; Elmore & Pohlmann, 1978; Gage, 1965; and Ryans, 1960).

Although many attempts have been made to determine a conception of effective teaching and strive to identify specific attributes and instructional classroom skills that would contribute to teacher effectiveness, it has been noted that there is no generic element for teacher effectiveness. Teacher effectiveness has been examined in relation to professional characteristics such as educational attainment, professorial ranks, and years of teaching experience as well as certain personality characteristics, but not much empirical evidence dealt on specific characteristics of students, teachers and the course itself that have a possible influence on teacher effectiveness (Travers & Rebore, 1995; Gonzales in Limson & Cacho, 1994).

The purpose of this research is to investigate the relationship of course characteristics, student characteristics and teacher characteristics with student ratings on teacher effectiveness. Likewise, it sought to determine which among these specific characteristics are predictive of teacher effectiveness.

#### Method

The correlational research design was used to determine a probable relationship of teacher, student and course characteristics with student ratings on teacher effectiveness. An attempt was likewise made to predict the criterion variable (teacher effectiveness) from the knowledge of 6 independent variables.

The Student Instructional Report (SIR) of John A. Centra and F. Reid Creech (1976) was used for the study. The SIR is a 38-item instrument designed to assess seven aspects of teacher effectiveness, namely; communication, planning and organization, knowledge of the subject matter, teaching skills, faculty and student interaction, tests and exams, and classroom management. It also consists of descriptive items on the perceived level of course difficulty, student's level of interest and expected grade, teachers' perceived strictness-leniency, teaching methodology and pace in teaching. Analysis of the SIR indicates a reliability coefficient which ranges from 0.73 - 0.94. This instrument was administered to a sample of 400 classes drawn from a community college in Metro Manila.

Philippine Journal of Educational Measurement • Volume 10

The data obtained from the SIR were decoded, tallied and analyzed using the following statistical tools: (a) mean scores and standard deviation, (b) Pearson product-moment correlation which would determine the degree of the relationship between students' ratings and course, student, and teacher characteristics; and (c) Stepwise regression which would estimate the value of the deendent variable or criterion from the knowledge of the other six independent variables or predictors; the regression equation would be the formula to predict the criterion.

## Results and Discussion

Responses on the student instructional report were tallied and statistically analyzed using descriptive and inferential statistics.

The summary of significant correlates of student ratings on teacher effectiveness and teacher characteristics is depicted in Table 1. From the table, it appears that teaching methodology obtained the highest correlation with a computed rvalue of - 0.420, then followed by strictness-leniency (r = +0.208) and pace in teaching (r = 0.095). Out of the three teacher characteristics, it is apparent that only two correlated significantly at 0.01 probability level. Perceived strictnessleniency refers to students' description of their teacher in so far as implementing rules and requirements in class.

Table 1.	Correlation	coefficients	between	student	ratings	and	teacher
	characteristi	cs					

Predictors	Obtained r	Significance
Perceived Strictness-Leniency	+0.208	Highly Significant **
Teaching Methodology	-0.420	Highly Significant **
Pace in Teaching	-0.095	Not Significant

\* .05 level = 0.096

\*\* .01 level = 0.127

From the results, it can be inferred that an unreasonably strict teacher who imposes rules and regulations which are excessive and very difficult to comply with will most likely receive low student ratings. On the other hand, a lenient teacher who cannot enforce class rules and requirements consistently and has the tendency to accommodate all students' request are likely to obtain high evaluation ratings.

Teaching methodology refers to the variability in the utilization of a particular teaching approach. Results reveal a negative relationship between teaching methodology and student ratings. This may imply that a teacher who always vary his/her teaching methodology will most likely obtain higher student evaluations compared to a teacher who uses the same teaching methodology all throughout a given period. Further analysis showed that pace in teaching obtained a computed r value of -0.095 which did not exceed the critical r value of 0.096. Hence, the null hypothesis is not rejected which declares that there is no significant relationship between pace in teaching and student ratings on teacher effectiveness.

As depicted in Table 2, student characteristics such as level of interest and expected grade obtained computed r values of +0.313 and -0.257, respectively. The computed r values of +0.313 and -0.257 exceeded the tabular r value of 0.196 at 0.01 probability level. A positive relationship between level of interest and student ratings may connote that a student who is very much interested in the course will respond to the teaching evaluation positively, thus, giving the teacher a high evaluation rating. On the other hand, a student who seems to be uninterested in the subject matter will tend to give the teacher a low evaluation rating. Moreover, a student who expects a high grade in the course/subject will most likely give high evaluation ratings compared to a student who expects a low grade.

1016.5	Predictors	Obtained r	Significance
0.00	Level of Interest	+0.313	Highly Significant **
	Expected Grade	-0.257	Highly Significant **

Table 2. Correlation coefficients between student ratings and student characteristics

\* .05 level = 0.096

\*\* .01 level = 0.127

Table 3 shows the correlation coefficient between perceived level of course difficulty and student ratings. Although, the computed r value of -0.097, which is considered negligible, was obtained, further analysis showed that the r value exceeded the tabular r value of 0.0961. Hence, the null hypothesis that there is no significant relationship between student ratings and level of course difficulty is rejected (p<.05). This finding may suggest that a student who perceives the subject/course to be somewhat elementary or easy for his/her ability, tends to rate his/her teacher at a very favorable light. However, if the subject/course is perceived to be very difficult, there is the inclination to rate the teacher low.

Philippine Journal of Educational Measurement • Volume 10

Table 3.	Correlation	coefficient	between	student	ratings	and	course
	characteristics						

	Predictors	Obtained r	Verbal Interpretation
ain a	Level of Difficulty	-0.097	Significant **

#### \* 05 level = 0.096

\*\* 01 level = 0.127

The significant variables that emerged were selected and expressed as regression equation to predict the value of the dependent or criterion variable. The three elements of teacher characteristics and the three other variables for student and course characteristics were considered the independent variables/ predictors, with student ratings as the criterion. The summary of the multi-steps procedure in predicting student ratings on teacher effectiveness is delineated in Table 4. Four out of the six independent variables were selected to be the best set of predictors which explained 29.34% of the variance in the student ratings  $(MR^2 = 0.293389)$ . The highest significant correlate was teaching methodology which showed 17.63% effect on the criterion. This was followed by students' level of interest which presented an additional 7.21%. Perceived strictness-leniency and level of course difficulty produced an increase of 3.58% and 0.92%, respectively. The regression equation for the predicted student ratings on teacher effectiveness is equal to + 3.605793 - 0.3235412 (teaching methodology) +0.3104172 (level of interest) +0.1989687 (strictness-leniency) -0.1117434 (level of course difficulty).

Hence, the null hypothesis which states that there are no significant relationships between student ratings as the dependent or criterion variable and the specific independent variables of teaching methodology, level of interest, perceived strictness-leniency, and level of course difficulty is rejected.

Variable Added	MR	MR <sup>2</sup>	
Teaching Methodology	0.419894	0.176311	
Level of Interes	0.498368	0.248370	
Perceived Strictness-Leniency	0.533077	0.284171	
Perceived Level of Course Difficulty	0.541654	0.293389	

Table 4. Summary table for stepwise regression

#### General Discussion

In this research, we investigated the relationship of student ratings on teacher effectiveness with course, student and teacher characteristics. At the theoretical level, the significance of the study is two- fold. First, it provides empirical support that student ratings on teacher effectiveness could be influenced by so many variables that are difficult to control. This position has been confirmed by reports that, although student ratings in general indicate student judgment about what the teacher does in the course and how the teacher performs, uncontrollable variables such as specific characteristics of students, teachers, or courses that have little to do with actual little performance of effectiveness come into play (Centra & Creech, 1976; Travers and Rebore, 1995).

Another theoretical contribution of this study relates to the issue of validity of the evaluation results. There are different positions regarding this issue. Evaluation results, generally, provide a basis upon which administrative decisions can be made. Such decisions may cover promotions in rank, reclassification to another faculty level, increase in salary, and tenure. Likewise, evaluation results serve as a feedback mechanism for the teacher's self and professional improvement (Gage, 1967). Johnson (1967), in addition, recognizes the importance of evaluation made by the students. He further asserts that:

It is often the individual student who knows best whether or not he is learning. It is the student who knows whether a course is stimulating to him to learn or he cannot understand or already knows what is being discussed. It is the student who can best formulate those fundamental and personal questions so bothering him that he cannot really proceed to other academic matters. It is the student who can best evaluate when he is beginning to integrate the process of learning with the problems he continually confronts in life.

On the other hand, some studies claim the passive subjectivity of teacher evaluation (Gustad,1967). Since teacher evaluation is regularly conducted, students who are answering the form could have formed a response set (Lapena, 1987). This may also suggest that students simply fill out the form without even reading the items. In extreme cases, students could be familiar with the evaluation instrument that they do not treat it objectively.

The results of the study clearly showed that the teacher's teaching methodology and perceived strictness-leniency had a significant relationship with student ratings. It was observed that teaching methodology is

Philippine Journal of Educational Measurement • Volume 10

moderately correlated with student ratings with a computed r value of -0.420. In relation to this finding, it is possible to infer that methods used in instruction have a bearing on how students' perceive teacher effectiveness. Centra and Creech (1976), in their study, revealed that teachers who often use the lecture method is evaluated significantly lower than teachers who utilize group discussions. In the group discusion approach, a teacher may ask questions, listen to student answers and insights, react and probe deeper into the information given. Tharp and Gallimore (1991) identified several advantages of group discussions. According to them, students are directly involved and are given the chance to participate. Likewise, it provides the students the opportunity to express themselves, justify their opinions, come up with their own stand on a particular issue, and assume responsibility for their perceptions. In effect, this minimizes the traditional role of the teacher as the sole provider of knowledge and optimizes critical thinking and student involvement. Reports also showed that teachers who employ discussion groups in their classes are rated higher because of the opportunity it gives to the students to interact with one another (Centra & Creech, 1976). Kulik, Kulik, and Carmichael (1974, in Centra and Creech) mentioned that teachers who use individualized instructional methods/approaches are rated higher than plain lecture. From the literature cited, it can be hypothesized that the way a teacher imparts knowledge to his/her students has an influence on how she/he will be evaluated. It can also be inferred that students tend to prefer teaching approaches which allow them to express their opinions, participate and pursue ideas that interest them.

Perceived teacher's strictness-leniency obtained a computed r value of +0.208 which indicates a significant relationships with student ratings. From this finding, it seems that the manner a teacher implements rules and requirements has an influence on how he/she is evaluated by his/ her students. Peck and Tucker (1970) suggested that teachers should establish a relaxed and conducive atmosphere in class. This is characterized by fewer directions, less criticism and authority, and negative feedback. Likewise, teachers should possess warmth democratic attitudes especially when enforcing rules and regulations in class (Hamcheck in Travers and Rebore, 1995). It is, therefore, essential that teachers explain the rationale as well as the importance why such rules and regulations are being implemented. If this is done correctly and properly, students will most likely know the expectations of their teachers.

Results further showed that students' level of interest, expected grade, and perceived level of course difficulty had a significant relationship with student ratings. One possible explanation for the significant loading of level of interest is that students who have more interest in the subject matter has the tendency to rate the teacher as more effective. Another possibility can be taken from the end point of the teacher. A teacher, for example, who shows less interest and enthusiasm in teaching a course/subject may contribute to his/her own evaluation rating. In relation to this, Collins (in Hamachek, 1989) described enthusiasm as one of the notable characteristics among effective teachers. He further added that enthusiastic teachers bring vitality and energy into classroom. This according to Collins, is typified by certain behavioral qualities such as liveliness, fervor, and activity. Moreover, researchers seem to concur that one characteristic considered important by students for effective teaching is the ability of the teacher to stimulate students'interest (Marsh, 1983, 1980; Marques, Todd, Lane, & Dorfman, 1979; Gadzella, 1971; Musella & Rusch, 1968; Downie, 1952; and Smith, 1944).

The variable students' expected grade obtained a computed r value -0.257 which indicates that it is significantly correlated with student ratings. Holmes (in Centra & Creech, 1976) reported that students tend to evaluate their teachers favorably or unfavorably depending on the grades students anticipate to receive. For example, if a student obtained a grade lower than he/she expected, he/she will most likely give his/her teacher a low evaluation rating. This observation likewise confirms the findings of several empirical studies conducted in this area (Marsh, 1983, 1980; Gonzales, 1991 in Cacho & Limson, 1994). However, some studies have found no evidence of a large grading standard effects on student ratings (Abrami, 1980; Centra & Creech, 1976).

Apparently, it seems that students who find a subject difficult would tend to rate a teacher handling the subject less favorably than a teacher handling a subject which is perceived to be less difficult. In relation to this finding, students may have the perception that the level of course difficulty is defined by the number of requirements assigned by the teacher. Although, this can be considered as a valid point on the part of the students, teachers should make an effort to explain why requirements are necessary for the subject as well as his/ her expectations of the students.

Based on this discussion, the practical implications of the results became evident. Now that we are aware of the possible influence of specific teacher, student and course characteristics to student ratings of teacher effectiveness, possible interventions may be implemented to neutralize its effects. For example, since it became evident that teaching methodology obtained the highest correlation coefficient and contributed significantly to the prediction of student ratings, primary importance should be given to teacher training, especially in skills acquisition/enhancement (e.g. use of various teaching and instructional strategies, multi-media presentations, innovative teaching tech-

Philippine Journal of Educational Measurement • Volume 10

niques, etc.). Teachers likewise should feel free to experiment on possible teaching strategies which they feel are effective and contributory to students' learning. On the other hand, students should be informed and oriented about the value of teacher evaluation. They should be made aware that teacher evaluation should be handled constructively and objectively. This move, hopefully, may make them feel that their comments and feedback are given importance.

On the whole, this study recognized that certain specific course, teacher and student characteristics have an influence on student ratings. We should realize, however, that the relationship between these variables and student ratings is not yet the entire picture. Research efforts should still be continued so as to unravel other variables, based on literature or by mere intuition, that can possibly influence student ratings. Although, Philippine universities and colleges, in general, put a high premium as far as quality education is concerned, what should be remembered, however, is not so much what the teacher teaches but how he/she teaches. Or in other words, students would prefer teachers who express rather than teachers who impress. To end, allow me to cite Anthony de Mello, an Indian Jesuit priest, who said something about student-teacher relationship, and I quote, "Education is not preparation for life. It should be life".

#### References

- Abrami, P. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? *Journal of Educational Psychology*, 72, 107-118.
- Cacho, L. & Limson, N. (1994). The relationship of teacher characteristics and student rating of teacher effectiveness. Unpublished undergraduate Thesis: De La Salle University.
- Caupayan, R. (1980).Teaching effectiveness as perceived by students, teachers and administrators of the Secondary Schools directed by the Sisters of St. Columban in Northwestern Mindanao. Unpublished masteral Thesis: De La Salle University.
- Centra, J. A. (1973). The effectiveness of student feedback in modifying college instruction. *Journal of Educational Psychology*, 65, 395-401
- Centra, J. & Creech, R. (1976). Two studies on the validity of the student instructional report. Educational Testing Service: New Jersey.
- Coffman, W. E. (1954) Determining students' concept of effective teaching from their instructors, *Journal of Education Psychology*, 45, pp. 277-286
- Crawford, P. L. & Bradshaw, H. L. Perception of characteristics of effective university teachers: A scaling analysis in student ratings: Reliability, validity and usefulness. *Review of Educational Research*, 41

Darshan, S. (1976) Ratings by high and low achieving students of teaching qalities of college professors. *Psychological Reports*, 38, 939-942.

Elmore, P.B. & Pohlmann, J. T. (1978). Effect of teacher, student, and class characteristics on the evaluation of college instructors. *Journal of Educational Psychology*, 7, 187-192.

- Endle, S., Murray, H. G., & Rushton, J. P. (1985). Journal of Educational Psychology, 77, 394-407
- French, G. M. (1957). College students' concept of effective teaching determined by an analysis if teacher ratings as found in Student ratings: Reliability, validity and importance. *Review of Educational Research*, 41, p. 515.
- Gadzella, B. M. (1971). College students views and ratings of an ideal professor. Review of Educational Research, 41, 515.
- Gage, N. L. (1961). The appraisal of college teaching. Journal of Higher Education, 32, 17-22.
- Gage, N. L.(1972). Teacher Effectiveness and Teacher Education. Pacific Book Publishers: California.
- Gustad. J. W. (1967) Evaluation of Teaching Performance: Issues and Possibilities. American Council of Education.
- Guthrie, E. R. (1949). The evaluation of teaching in student ratings: Reliability, validity and importance. Review of Educational Research, 41, p.515.
- Hamchek, D. (1989). Psychology in Teaching, Learning and Growth. Allyn and Bacon: U.S.A.
- Johnson, J. (1967). Instruction: From the Consumer's View. American Council of Education.
- Marques, T.E., Lane, D. M., & Dorfman, P. W. Toward the development of a system for instructional evaluation: Is there consensus regarding what constitutes effective teaching? *Journal of Educational Psychology*, 71, 840-849.
- Marsh, H.W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic setting and their relation to student/ course/instruction characteristics. Journal of Educational Psychology, 75, 150-166.
- Marsh, H. W. (1980). The influence of student, course, and instructor characteristics in evaluation of university teaching. *American Educational Research Journal*, 17, 219-237
- Meredith, G. M. (1988). shared components in students' ratings of instructors' effectiveness and need for improvement. *Psychological Reports*, 62, 832-834.
- Musella, D. & Rusch, R. Student opinion on college teaching. Improving colleges and university teaching, 16, pp 137-140.
- Smith, A.A. (1944). College teacher evaluation by students. Sociology and Social Research, 28, 471-478.

Philippine Journal of Educational Measurement • Volume 10

Tharp, R.G., & Gallimore, I.R. (1991). The instructional conversation: Teaching and Learning in social activity. Washington, DC: National Center for Research on Cultural Diversity and Second Lange Learning.

Travers, P.D. & Rebore. R. W. (1995). Foundations of Education. Allyn and Bacon: U.S.A.

- Steffens, P.J. (1990). Characteristics of teacher efficacy as perceived by teachers of the year. *Dissertation Abstracts International*, 51-52, 4094.
- Waters, M., Kemp, E., & Pucci, A. (1988). High and low faculty evaluations; Descriptions by students. *Teaching Psychology*, 15, 203-204.
- Wilen, W. (1990). Forms and phases of discussion. In W. Wilen (Ed.), Teaching and learning through discussion (pp 3-24). Springfield, IL: Charles C. Thomas.

Woolfolk, A. E. (1995). Educational Psychology. Allyn and Bacon: U.S.A.

A STATEMENT AND A STATEMENT AN



# **Center for Educational Measurement, Inc.**

24th Floor, Cityland Pasong Tamo Tower 2210 Chino Roces Ave., Makati City

Copyright 1999