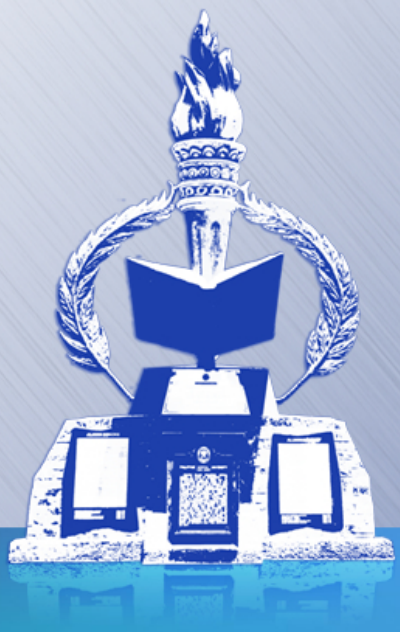
March 2013

Development and Evaluation of a Critical Thinking Test in Mathematics for Grade III Pupils

Esperanza Magbujos Batitis
Philippine Normal University



DEVELOPMENT AND EVALUATION OF A CRITICAL THINKING TEST IN MATHEMATICS FOR GRADE III PUPILS

A Seminar Paper Presented
to the Faculty of the College of Science
Philippine Normal University
The National Centre for Teacher Education
Taft Avenue,Manila

In Partial Fulfillment of the
Requirements for the Degree
Master of Arts in Teaching Mathematics

---

by
Esperanza Magbujos Batitis
March 2013

Philippine Normal University
National Center for Teacher Education

APPROVAL SHEET

This special project entitled **DEVELOPMENT AND EVALUATION OF A CRITICAL THINKING TEST IN MATHEMATICS FOR GRADE III PUPILS** prepared and submitted by **ESPERANZA MAGBUJOS BATITIS** in partial fulfillment of the requirements for the degree Master of Arts in Teaching Mathematics, is hereby recommended for acceptance.

4/12/2013
Date

ELISA S. BACCAY, Ed. D.
Adviser

Accepted in partial fulfillment of the requirements for the degree Master of Arts in Teaching Mathematics.

5/24/13
Date

GLADYS C. NIEVERA, Ph. D.
Head, Department of Mathematics

5/25/13
Date

ROSEMARIEVIC V. DIAZ, Ph. D.
Dean, College of Science

ACKNOWLEDGEMENT

The researcher wishes to express her profound gratitude to the Lord Almighty for showering her with, his blessings and grace making this study a reality.

Heartfelt appreciation is likewise conveyed to the people who extended to the researcher their valuable encouragement, assistance and moral support in the completion of this study. Special mention is accorded to the following persons;

Dr. Elisa S. Baccay, Ed. D. in Educational Administration – Math Department P.N.U. Manila, the researcher adviser for her intelligent guidance and professional help.

Dr. Rene R. Belecina, Ph. D. – Math Department P.N.U. Manila, Miss Severa Salamat – EPS in Mathematics Division of Santa Rosa City for validation, Dr. Peter Obias – Psychology Department P.N.U. Manila, Mrs. Ana Maria S. Macabata, Ms. Ammie T. Dy Pico and Mrs. Angelita M. Batitis Teachers of Santa Rosa City for evaluating my instrument .

To our dearest panelists, Professor Allan Reyes, Professor Myla Esperanza, Dr. Rene Belecina for giving valuable comments and suggestions which led to further improvements of the study.

Mrs. Zenaida Bernardino, Ms. Josie Bela-ong, Ms. Fey delos Santos,Mrs. Virgina Naorbe, Mrs. Evezer Giron, Mrs. Maricel Cabico, Mrs. Mary Anne Abelida,Mrs. Lina Balbieran, Ms. Bibiana Ferrer, Mrs. Teresita Beato, and all other Grade I teachers of SRES Central – I whose inspiration and moral support guided her in the pursuit of this work. Mrs. Richelda Merciales, Mrs. Victoria Mariñas and their respective pupils who were used as respondents for their cooperation and sincerity in taking the test during the try-outs.

Ms. Hershey Joy Batitis for her assistance in the use of computers and helping the researcher typing the text.

Mrs. Leonila M. Buitizon, Principal IV of Santa Rosa Elementary School Central – I for permitting the researcher to gather data from her school teachers and pupils.

To all these people go her sincerest thanks and fondest hopes that they succeed in their chosen fields. May the Lord bless them always.

E.M.B.

D E D I C A T I O N

This work is lovingly dedicated to my parents,
Phievie, Hazzell, Hershey Joy, Hensly Paul and Honey Grace
whose love, concern and encouragement have never ceased
to lift me up.

ABSTRACT

Statement of the Purpose

This study sought to develop a critical thinking test instrument for Grade III pupils of Santa Rosa Elementary School Central – I, Santa Rosa City, Laguna.

Specifically, it aimed to do the following:

1. Construct test items that will measure critical thinking skills in Mathematics among Grade III pupils in terms of

    1.1. defining and clarifying the problem

    1.2 judging information related to the problem

    1.3 solving problems/drawing conclusions

2. Evaluate the developed test in terms of content and

3. Estimate the reability of the test using test-retest method.

Procedures

The study adopted the descriptive method of research. Data were obtained from eighty Grade III pupils of Santa

Rosa Elementary School Central – I of Santa Rosa City, Laguna. The developed test went through five steps. Step 1 dealt with preliminary activities which includes gathering relevant information on critical thinking and test development, planning the test, making the table of specification, and selecting the appropriate test format. Step 2 included writing the preliminary form of the test. The test items were based on the purpose of the test, identifying the relevant constructs, review and revision of the adviser, rewriting the test, incorporating the comments and suggestions for the improvement of the test, administering the revised preliminary form of the test. Step 3 asking experts for their comments, suggestions and recommendations for the improvement of the test. Step 4 dealt with the evaluation and validation of the test experts in terms of content. Step 5 covered the evaluation of the test which included the estimation of the reliability of the test using the test-retest method.

The descriptive statistics of mean, frequency, percentage and correlation (Appendix B) were used in the analysis of data.

Findings

The following were the findings of the study:

1. The evaluators *strongly agree* that the developed test items truly measures the critical thinking skills of Grade III pupils.

2. The developed test has a high reliability index of .73. Since the test scores were correlated, this is another indicator of the validity of the test.

Conclusions

In light of the findings of the study, these conclusions were made:

1. The developed critical thinking skills test is valid and reliable.
2. The developed test is usable and may be used to measure the critical thinking skills of Grade III pupils.

Recommendations

In light of the findings and conclusions of the study, the following recommendations are given

1. Subject the developed instrument to further validity and reliability.

2. The instrument can be used to determine the level of the critical thinking skills of Grade III pupils and the result may be used as basis in preparing appropriate learning activities in Mathematics that will enhance their critical thinking ability.

3. To develop other critical thinking tests appropriate for other grade levels.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

More recently, the partnership for 21st century skills has identified critical thinking skills as one of several learning and innovation skills necessary for post – secondary education and the workforce. In addition, the core state standards reflect critical thinking as a cross – disciplinary skill vital for college and employment. Critical reasoning is deemed valuable for people of all ages.

Learning to think critically in Math is the foundation on which individuals may developed meaningful mathematical study. Every discipline can be understood only through thinking. We know mathematics, not when we can recite mathematical formula, but when we can think mathematically.

A critical thinking test reveals strengths, weaknesses and levels of thinking (Facione P.A. 2011). Answering a critical thinking test instrument helps us to a higher level thinking, sometimes called the scientific way of thinking because it involves analyzing and evaluating facts. Work on improving pupils critical thinking skills helps them make more intelligent decisions.

Theoretical/Conceptual Paradigm

In 1910, Dewey said that thinking does not occur spontaneously but must be evoked by problems and questions or by some perplexity, confusion or doubt. Further, he said that observations or data at hand cannot supply the solution. They can only suggest it. It is this demand for the solution that steadies and guides the entire process of critical thinking the nature of the problem fixes the end of thought and the end controls the process of thinking.

Papert, in 1980, claims that

> *"Critical thinking is at the heart of effective reading, writing, speaking and listening. It enables us to link together mastery of context with such diverse goals as self-esteem, self-discipline, multi-cultural education effective cooperative learning, and problem solving. It enables all instructors and administrators to raise the level of their own teaching and thinking".*

The Asian Polymer Association Delphi report recommends that "from early childhood, people should be taught, for example, to reason, to seek relevant facts, to consider options, and to understand the views of others" (Facione, 1990). Moreover, the report maintains that explicit instruction dedicated to critical thinking skills, abilities, and dispositions should be built into all levels

of the K-12 curriculum, rather than being limited to junior high or high school students.

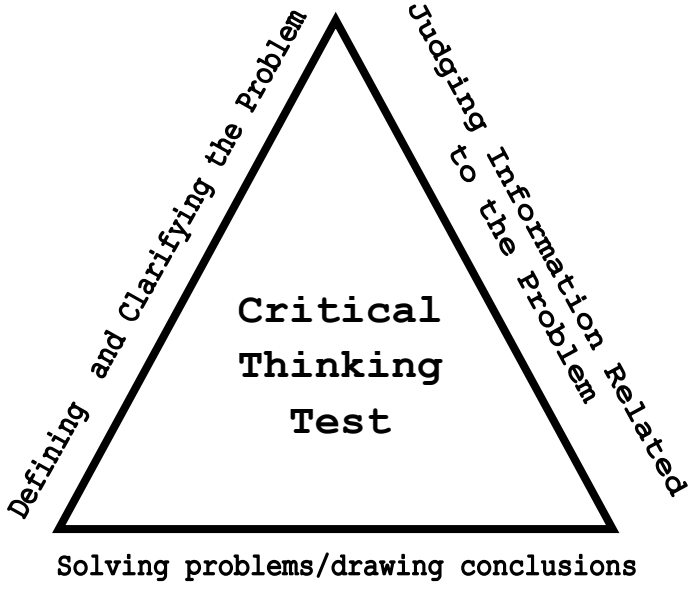


Figure 1 Conceptual Paradigm

This figure served as the principal guide in this study, following Kneedler's view of critical thinking. The components of the developed critical thinking test are shown at the sides of the triangle. These components are; A. Identifying central issues or problems, comparing similarities and differences,determining which information is relevant and formulating appropriate questions. B. is determined through consistency and C. is determined through recognizing the adequacy of data and through predicting probable consequences.

To help teacher incorporate critical thinking in the classroom is to help them asking questions that call for analysis, systhesis and evaluation.

Statement of Purpose

This study sought to develop a critical thinking test for Grade III pupils of Santa Rosa Elementary School Central – I, Santa Rosa City, Laguna.

Specifically, it aimed to do the following;

1. Construct test items that will measure critical thinking skills in mathematics among grade III pupils with the following components

    1.1 Defining and Clarifying the Problem

    1.2 Judging Information Related to the Problem

    1.3 Solving Problems/Drawing Conclusions

2. Evaluate the developed test and

3. Estimate the reability of the test using test-retest method.

Significance of the Study

This study is deemed important as it provides a framework which can serve as a model for test construction that may be useful to researchers and teachers. It provides on going mathematical assessment data that can be used to make instructional decisions and help students succeed

Critical thinking is at the heart of effective reading, writing, speaking and listening. It enables us to link together mastery of context with such diverse goals as self-esteem, self-discipline, multi-cultural education effective cooperative learning, and problem solving. Teaching through problem solving is fun and when learning is fun pupils have a better chance of remembering it later and it enables all instructors and administrators to raise the level of their own teaching and thinking.

Teachers try new methods, invent new ways of illustrating new ideas, try several approaches to bring forth the best from pupils. Concerned as they are, they strive to exert a positive influence on the forces that affect education. Given research in cognitive psychology, some educators believe that schools should focus on teaching their pupils critical thinking skills and

cultivation of intellectual traits. In this study, the teacher will be guided properly on how to achieve the goal of developing a critical thinking test instruments as one process of cultivating intellectual traits of pupils.

Finally, this research provides a better understanding of the importance and the impact of the critical thinking test instruments on pupils learning and mathematics achievement.

The findings revealed in this study may prompt concerned school authorities to improved tests focused on critical thinking.

Scope and Delimitation of the Study

This study was conducted in Santa Rosa Elementary School Central-I, Santa Rosa City, Laguna during the school year 2012-2013. The respondents were the Grade III pupils from the Fast Learner Group and Section II.

In developing the test, the criteria were culled from Kneedler's list (1985).

The developed test was evaluated by experts using validated questionaire (Appendix C).

The reliability of the test was estimated using the test-retest method.(Shuttleworth,2009)

Definition of Terms

The following terms are conceptually and operationally defined for a better understanding of the study.

- **Critical Thinking.** It is the ability to use many different processes of information to come to unique solutions to problems or thoughts / innovations. (Askville, 2012)

- **Critical Thinking Test.** It is the developed test this study that is composed of defining and clarifying the problem, judging, information related to the problem, solving problems/drawing conclusions. (Kneedler 1985).

- **Evaluation.** This is a process which examines student responses to individual test items in order to assess the quality of those items and of the test as a whole (Office of Educational Assessment University of Washington, 2009).

- **Reliability.** This refers to the consistency of scores or answers provided by an instrument (Fraenkel and Wallen,2008).

- **Table of Specifications.** It is a plan, a guide, or a blueprint to define as clearly as possible the scope and emphasis of the test, and to relate the objectives to the content (University of Kansas , 2009.)

- **Test.** It is an assessment intended to measure a test-takers knowledge, skills, aptitude and physical fitness. (wikipedia, the free encyclopedia).

- **Validity.** This refers to the appropriateness, meaningfulness, correctness, and usefulness of any inferencess a researcher draws based on the data obtained through the use of a instrument (Fraenkel and Wallen,2008)

Chapter I

THE PROBLEM AND ITS BACKGROUND

Introduction

Critical thinking skills are useful in real life situations. Critical thinking is an approach to thinking in which a person visualizes an idea and then goes about the tasks of taking the steps necessary to reach a conclusion (AERA,1999). It involves research, investigation, evaluation, conjecture and implementing. Having critical thinking ability is vital to many professions in today's age of information society.

It is a teacher's responsibilty to foster the ability to think critically among pupils. In the classroom pupils must be taught to look at things critically and to evaluate parts of a discussion and life situations. Teaching them to recognize the strengths and weaknesses of arguments will help them make better choices for themselves both in and out of the classroom.

Educators have long been aware of the importance of critical thinking skills as an outcome of student learning.

Chapter II

REVIEW OF RELATED LITERATURE

In this chapter, the researcher presents the related concepts formulated by well-known educators psychologist and researchers in the field about the development and evaluation of a critical thinking test in Mathematics. Their studies is considered as the bases of answering the needs of the study.

Critical Thinking

McPeck (1981) defines critical thinking as the propensity and skill to engage in an activity with reflective skepticism. Reflective and reasonable thinking that is focused on deciding what to believe or do while Lipman (1988) defines it as skillful, responsible thinking that facilitates good judgment because it 1) relies upon criteria, 2) is self-correcting, and 3) is sensitive to context.

As stated by Paul (1992), critical thinking is disciplined, self-directed thinking that exemplifies the perfections of thinking appropriate to a particular mode or domain of thought, thinking that is goal-directed and purposive.

The Cognitive Psychological Approach.

The cognitive psychological approach contrasts with the philosophical perspective in two ways. First, cognitive psychologists, particularly those immersed in the behaviorist tradition and the experimental research paradigm, tend to focus on how people actually think versus how they could or should think under ideal conditions (Sternberg, 1986). Second, rather than defining critical thinking by pointing to characteristics of the ideal critical thinker or enumerating criteria or standards of "good" thought, those working in cognitive psychology tend to define critical thinking by the types of actions or behaviors critical thinkers can do. Typically, this approach to defining critical thinking includes a list of skills or procedures performed by critical thinkers (Lewis & Smith, 1993).

Philosophers often criticized aspect of the cognitive psychological approach as being reductionist—reducing a complex orchestration of knowledge and skills into a collection of disconnected steps or procedures (Sternberg, 1986). For example, Bailin (2002) argues that it is a fundamental misconception to view critical thinking as a

series of discrete steps or skills, and that this misconception stems from the behaviorist's need to define constructs in ways that are directly observable. According to this argument, because the actual process of thought is unobservable, cognitive psychologists have tended to focus on the products of such thought—behaviors or overt skills (e.g., analysis, interpretation, formulating good questions).

Other philosophers have also cautioned against confusing the activity of critical thinking with its component skills (Facione, 1990), arguing that critical thinking is more than simply the sum of its parts (Van Gelder, 2005). Indeed, a few proponents of the philosophical tradition have pointed out that it is possible to simply "go through the motions," or proceed through the "steps" of critical thinking without actually engaging in critical thought (Bailin, 2002).

Definitions of critical thinking that have emerged from the cognitive psychological approach include the mental processes, strategies, and representations people use to solve problems, make decisions, and learn new concepts (Sternberg, 1986); the use of those cognitive

skills or strategies that increase the probability of a desirable outcome; and seeing both sides of an issue, being open to new evidence that disconfirms your ideas, reasoning dispassionately, demanding that claims be backed by evidence, deducing and inferring conclusions from available facts, solving problems, and so forth" (Willingham, 2007).

The Educational Approach

Finally, those working in the field of education have also participated in discussions about critical thinking. Bloom (1956) and his associates are included in this category. Their taxonomy for information processing skills (1956) is one of the most widely cited sources for educational practitioners when it comes to teaching and assessing higher-order thinking skills. Bloom's taxonomy is hierarchical, with *comprehension* at the bottom and *evaluation* at the top. The three highest levels, consisting of analysis, synthesis, and evaluation are frequently said to represent critical thinking (Kennedy et al., 1991). The educational approach is based on years of classroom experience and observations of student learning, unlike both the philosophical and the psychological traditions

(Sternberg, 1986). However, some have noted that the educational approach is limited in its vagueness. Concepts within the taxonomy lack the clarity necessary to guide instruction and assessment in a useful way (Ennis, 1985; Sternberg, 1986). Furthermore, the frameworks developed in education have not been tested as vigorously as those developed within either philosophy or psychology (Sternberg, 1986).

Areas for Agreement

Despite differences among the three schools of thought and their approaches to defining critical thinking, there exist areas for agreement. First, researchers of critical thinking typically agree on the specific abilities encompassed by the definition, which include analyzing arguments, claims, or evidence (Ennis, 1985; Facione, 1990; Halpern, 1998; Paul, 1992); making inferences using inductive or deductive reasoning (Ennis, 1985; Facione, 1990; Paul, 1992; Willingham, 2007); judging or evaluating (Case, 2005; Ennis, 1985; Facione, 1990; Lipman, 1988; Tindal & Nolet, 1995); and making decisions or solving problems (Ennis, 1985; Halpern 1998; Willingham, 2007). Other abilities or behaviors identified as relevant to critical thinking include asking and answering questions for clarification (Ennis,

1985); defining terms (Ennis, 1985); identifying assumptions (Ennis, 1985; Paul, 1992); interpreting and explaining (Facione, 1990); reasoning verbally, especially in relation to concepts of likelihood and uncertainty (Halpern, 1998); predicting (Tindal & Nolet, 1995); and seeing both sides of an issue (Willingham, 2007).

Dispositions

Most researchers also agree that in addition to skills or abilities, critical thinking also involves dispositions (Facione, 1990). As early as 1985, researchers working in the area of critical thinking recognized that the *ability* to think critically is distinct from the *disposition* to do so (Ennis, 1985). Empirical evidence appears to confirm the notion that critical thinking abilities and dispositions are, in fact, separate entities (Facione, 2000). These dispositions have variously been cast as attitudes or habits of mind. Facione (2000) defines critical thinking dispositions as "consistent internal motivations to act toward or respond to persons, events, or circumstances in habitual, yet potentially malleable ways". Researchers tend

to identify similar sets of dispositions as relevant to critical thinking. The most commonly cited critical thinking dispositions include open-mindedness (Bailin et al., 1999; Ennis, 1985; Facione 1990, 2000; Halpern, 1998); fair-mindedness (Bailin et al., 1999; Facione, 1990); the propensity to seek reason (Bailin et al., 1999; Ennis, 1985; Paul, 1992); inquisitiveness (Bailin et al., 1999; Facione, 1990, 2000); the desire to be well-informed (Ennis, 1985; Facione, 1990); flexibility (Facione, 1990; Halpern, 1998); and respect for, and willingness to entertain, others' viewpoints (Bailin et al., 1999; Facione, 1990).

Facione (1990) stated that: although the identification and analysis of critical thinking skills transcend, in significant ways, specific subjects or disciplines, learning and applying these skills in many contexts requires domain-specific knowledge. This domain-specific knowledge includes understanding methodological principles and competence to engage in norm-regulated practices that are at the core of reasonable judgments in those specific context. Too much of value is lost if critical thinking is conceived of simply as a list of logical operations and domain-specific knowledge is conceived of simply as an aggregation of information.

Development of Critical Thinking

Paul (1992) argues that typical school instruction does not encourage the development of higher-order thinking skills like critical thinking. Paul explains that knowledge is coterminous with thinking, especially good or critical thinking. However, typical school instruction, with its emphasis on the coverage of content, is designed as though recall were equivalent to knowledge. This type of lower-order learning is simply learning by rote or association, with the end result that students memorize material without understanding the logic of it. Students tend not to recognize that their assertions, beliefs, and statements have implications, and thus require evidence to support them. For most students, believing, not thinking, is knowing (Paul, 1992).

Despite evidence suggesting that the average person struggles to think critically, many researchers are sanguine about the capacity of humans to become critical thinkers with appropriate instruction. Kennedy et al. (1991) point out that empirical research suggests that students of all intellectual ability levels can benefit from critical thinking instruction. Similarly, Lewis and Smith (1993) argue that

critical thinking skills are for everyone, not just the gifted.

Critical Thinking in Children

Early research in the Piagetian tradition tended to view the cognitive processes of young children as being deficient in relation to those of older individuals. Many following this tradition interpret Piaget's stages of development to mean that young children are incapable of formal operations (abstract reasoning), which are required for critical thought (e.g., see summary in Kennedy et al., 1991). However, more recent research has found that young children engage in many of the same cognitive processes that adults do, concluding that there is a place for critical thinking in the lower elementary curriculum Gelman & Markman (1986). Silva (2008) argues that there is no single age when children are developmentally ready to learn more complex ways of thinking. Further, Willingham (2007) indicates that very young children have been observed thinking critically, whereas trained scientists occasionally fall prey to errors in reasoning. Kennedy, et al. (1991) surveyed the research literature and concluded that, although critical thinking ability appears to improve with age, even young children can benefit from

critical thinking instruction. The authors speculate that many of the earlier gloomy conclusions, vis-à-vis the limited critical thinking skills of young children, were spuriousdue to a lack of relevant background or content knowledge needed to engage in a task.

Bailin et al. (1999) argue that critical thinking instruction at the primary grade levels can include teaching students to value reason and truth; respect others during discussion; be open-minded; be willing to see things from another's perspective; perceive the difference between definitions and empirical statements; use cognitive strategies, such as asking for examples when something is unclear; and use principles of critical thinking, such as considering alternatives before making a decision.

Similarly, the APA Delphi report recommends that "from early childhood, people should be taught, to reason, to seek relevant facts, to consider options, and to understand the views of others" (Facione, 1990) Moreover, the report maintains that explicit instruction dedicated to critical thinking skills, abilities, and dispositions should be built into all levels of the K–12 curriculum, rather than being limited to junior high or high school students.

Empirical evidence supports the notion that young children are capable of thinking critically. Koenig and Harris (2005) have demonstrated that 3- and 4-year-old children will differentiate the credibility of various sources of information. In particular, 4-year-old children appeared to prefer the judgments of adult participants who had a history of being correct over those who were purposefully inaccurate. This finding was replicated in a number of other studies (Jaswal & Neely, 2006). Similarly, Lutz and Keil (2002) found that children as young as 4 years appeared to be aware that different people may possess differing domains of expertise and that these areas of expertise might be related to their credibility on certain topics. For example, a car mechanic's diagnosis of car trouble was found to be more credible than a doctor's. Finally, Heyman and Legare (2005) found that children between the ages of 7 and 10 became increasingly aware that people may have motives to distort the truth, whereas children younger than this were not consistently critical of the credibility of people with such motives.

The afore cited motions have serious instruction implications. This includes teachability and development of

critical thinking among learners . Fostering the development of critical thinking is considered vital in the intructional process.

The Teachability of Critical Thinking

Many critical thinking researchers maintain that critical thinking skills and abilities can be taught. Halpern (1998) offers evidence of two instructional programs aimed at improving the critical thinking skills and abilities of college students. In one study, students who were taught general problem-solving skills improved on Piagetian-inspired measures of cognitive development. In the other study, college students instructed in a specific type of problem-solving strategy produced mental math representations that were more like those of experts than of novices. In their review of the literature, Kennedy et al. (1991) concluded that instructional interventions aimed at improving students' critical thinking skills have generally shown positive results. In a meta-analysis of 117 empirical studies examining the impact of instructional interventions on students' critical thinking skills and dispositions, Abrami et al. (2008) found that these interventions, in general, have a positive impact, with a mean effect size of 0.34.

However, the distribution of effect sizes was highly homogeneous, with effect sizes varying dramatically by type of intervention and sample characteristics. Effect sizes for students in K–12 settings were higher than those observed among undergraduates.

Teaching for Transfer

As earlier noted researchers disagree on the extent to which critical thinking skills learned in one context are transferrable to new contexts, domains, and disciplines. Most researchers tend to agree, however, that transfer is unlikely to occur unless students are taught specifically to transfer. What does this mean from an instructional standpoint? First, students must be given opportunities to apply critical thinking skills and abilities in a wide range of contexts and subject areas. Second, instruction should emphasize executive functioning or metacognitive skills, such as setting goals, planning, and monitoring progress toward goals (Kennedy et al., 1991). Third, students should be sensitized to deep problem structure, because most students' thinking tends to focus on the surface structure of problems, or the superficial aspects of tasks (Halpern, 1998; Willingham, 2007). Hummel and Holyoak define structure sensitivity as the

ability to "code and manipulate relational knowledge" (as cited in Halpern, 1998). The goal of structure training is to enable students to recognize a particular problem structure whenever they see it—whether it appears in math, science, or social studies—so that they may deploy appropriate strategies. Structure training involves distributing practice in a variety of contexts and settings. Halpern points out that use of "authentic" or real-world learning activities helps to promote the transfer of critical thinking skills. Brown (1990) argues that domain-specific knowledge may also be necessary for young children to successfully transfer skills to new problems that display the same deep structure. She observes, "We conclude that even young children show insightful learning and transfer on the basis of deep structural principles, rather than mere reliance on salient perceptual features, when they have access to the requisite domain-specific knowledge to mediate that learning". Thus, teaching for transfer may also entail providing adequate instruction on relevant background information.

Teaching Strategies

Lessons involving higher order thinking skills require particular clarity of communication to reduce ambiguity and confusion and improve student attitudes about thinking tasks.

Lesson plans should include modeling of thinking skills, examples of applied thinking, and adaptations for diverse student needs. Scaffolding (giving students support at the beginning of a lesson and gradually requiring students to operate independently) helps students develop higher order learning skills. However, too much or too little support can hinder development.Useful learning strategies include rehearsal, elaboration, organization, and metacognition. Lessons should be specifically designed to teach specific learning strategies. Direct instruction (teacher-centered presentations of information) should be used sparingly. Presentations should be short (up to five minutes) and coupled with guided practice to teach subskills and knowledge.Teacher- and/or student-generated questions about dilemmas, novel problems, and novel approaches should elicit answers that have yet to be learned. Feedback providing immediate, specific, and corrective information should inform learners of their progress. Small group activities such as student discussions, peer tutoring, and cooperative learning can be effective in the development of thinking skills. Activities that involve challenging tasks, teacher encouragement to stay on task, and ongoing feedback about group progress may engender critical thinking.

Specific Instructional Strategies

A number of researchers have recommended using particular instructional strategies to encourage the development of critical thinking skills and abilities. These include explicit instruction, collaborative or cooperative learning, modeling, and constructivist techniques. Critical thinking skills and abilities are unlikely to develop in the absence of explicit instruction (Abrami et al., 2008; Case, 2005; Facione, 1990; Halpern, 1998; Paul, 1992). Facione (1990) points out that this explicit instruction should also attend to the dispositional or affective component of critical thinking.

Another method recommended by several critical thinking researchers is a collaborative or cooperative approach to instruction (Abrami et al., 2008; Bailin et al., 1991; Bonk & Smith, 1998; Heyman, 2008; Nelson, 1994; Paul, 1992; Thayer-Bacon, 2000). This recommendation appears to be rooted in Piagetian and Vygotskyian traditions that emphasize the value of social interactions for promoting cognitive development (as summarized in Dillenbourg et al., 1996). Piaget touted the instructional value of cognitive conflict for catalyzing growth, typically achieved by interacting with another person

at a higher developmental stage. Along similar views, Vygotsky identified the zone of proximal development as the distance between what an individual can accomplish alone and what he/she can accomplish with the help of a more capable other (either a peer or an adult). Each of these approaches highlights the potential for cognitive improvement when students interact with one another (as summarized in Dillenbourg et al., 1996).

Proponents of collaborative or cooperative learning include Thayer-Bacon (2000), who emphasizes the importance of students' relationships with others in developing critical thinking skills. Supporters also include Bailin et al. (1999), who argue that critical thinking involves the ability to respond constructively to others during group discussion, which implies interacting in pro-social ways by encouraging and respecting the contributions of others. Similarly, Heyman (2008) indicates that social experiences can shape children's reasoning about the credibility of claims. In their meta-analysis of 117 empirical studies on the effects of instructional interventions for improving students' critical thinking skills and dispositions, Abrami et al. (2008) found a small but positive and significant effect of collaborative learning approaches on critical thinking.

Nelson (1994) provides some clues as to how collaboration can prompt cognitive development among college students. According to Nelson, students' misconceptions interfere with their ability to acquire new knowledge, despite appropriate instruction. Collaborations create opportunities for disagreements and misconceptions to surface and to be corrected. Collaboration also provides a vehicle for students to attain necessary acculturation to the college learning environment and helps to make tacit disciplinary expectations more explicit for students.

He also points out that collaboration must be scaffolded, arguing that this scaffolding process has three stages. First, students must be prepared for collaboration by providing them with a common background on which to collaborate, such as common assigned readings. Second, student groups should be provided with questions or analytical frameworks that are more sophisticated than they would tend to use on their own. Finally, collaborative activities should be structured by specifying student roles and by creating incentives for all group members to actively participate. Bonk and Smith (1998) identify a number of classroom activities that build on the potential for collaboration to enhance learning. These activities include

think-pair-share,round-robin discussions, student interviews, roundtables, gallery walks, and "jigsawing."

In addition to explicit instruction and collaboration, several other strategies have been identified as helpful in promoting critical thinking. For example, teachers are urged to use constructivist learning methods, characterized as more student-centered than teacher-centered (Bonk & Smith, 1998; Paul, 1992). Constructivist instruction is less structured than traditional instruction, amplifying students' roles in their own learning and de-emphasizing the role of the teacher. Educators should model critical thinking in their own instruction by making their reasoning visible to students. This could be accomplished by "thinking aloud" so that students can observe the teacher using evidence and logic to support arguments and assertions (Facione, 2000; Paul, 1992). Specific classroom learning activities believed to promote critical thinking include the creation of graphic organizers, such as concept maps and argument diagrams (Bonk & Smith, 1998; Van Gelder, 2005); Know - Want to Know - Learned charts, which require students to identify what they already know about a topic, what they want to know, and what they have learned upon completing instruction; "in a nutshell" writings, which entail summaries of arguments; exit

slips, which identify the most important thing learned and the areas of needed clarity; problem-based learning, particularly the use of ill-structured problem contexts; and mock trials (Bonk & Smith, 1998).

Critical Thinking Assessment

Although critical thinking has often been urged as a goal of education throughout most of this century (Dewey 1910). And the Educational Policies Commission's The Central Purpose of American Education, 1961), not a great deal has been done about it. Since the early 1980s, however, attention to critical thinking instruction has increased significantly – with some spillover to critical thinking assessment, an area that has been neglected even more than critical thinking instruction. Critical thinking assessment requires that we be clear about what we are trying to assess.

Classroom and Statewide Assessments

The measurement of higher order thinking skills requires that students be unfamiliar with the questions or tasks they are asked to answer or perform and that they have sufficient prior knowledge to enable them to apply their higher order thinking skills to answer the question or solve the problem. Meeting these requirements is problematic in dealing with

either real-life or subject-matter domain problems. A recommended solution (Sugrue, 1994) is to examine students using multiple-item formats to measure multiple aspects of a domain. Student profiles of correct answers can help to determine the extent to which the overall assessment is novel for individual students.Psychological studies of transfer of higher order thinking skills suggest that skills taught in one context generalize to other contexts if explicit teaching for transfer and self-monitoring is included in the instruction. Other studies suggest that over long periods of time individuals, as a result of educational and other personal experiences, develop higher order skills (intellectual abilities) that apply to the solutions of a broad spectrum of complex problems. Conversely, psychometric studies of the generalizability of higher order thinking skills when measured by hands-on performance tasks show great method and task variability in student performance; that is, individual student achievement is highly dependent on the tasks and methods used in the assessment.

Estimates of the number of tasks needed in an assessment to achieve generalizable (reliable) results range from 8 to 23. Published tests of higher order thinking skills have been

found to be sensitive to instruction. Three studies of such tests involved single group pre- and posttest comparisons and two used quasi-experimental treatment and control groups. In all studies, results were statistically significant, but treatment effects were small. A recommended procedure (Ennis, 1993) for those who might want to use a published test to assess student levels of critical thinking or for evaluating programs or courses is to examine the test items for content validity, then to take the test themselves and examine their own performance critically.

Three item/task formats have been identified for use in assessing higher order thinking skills: (1) selection, which includes multiple-choice, matching, and rank order items; (2) generation, which includes short answer, essay, and performance items or tasks; and (3) explanation, which involves giving reasons for selection or generation responses. All three formats are valid for measuring some aspects of higher order thinking, but all three have some disadvantages. Multiple-choice and constructed-response items can be used for separately measuring many specific higher order skills such as deduction, inference, and prediction but are less successful for measuring synthesis and evaluation.

Some validity problems can be avoided by using answer justification, where subjects select responses and then justify in writing their reasons for the choices they made. Multiple-choice items can sample many aspects of the subject matter, and performance tasks and essays can deal with complex real-life problems that require students to employ a number of higher order skills in solving them. However, final solutions to performance tasks and essays may not reveal the application of many higher order skills. Performance tasks and essays usually sample only a few aspects of the subject matter. Portfolios are useful in promoting and informally assessing reflective thinking skills.Even though classroom teachers recognize the importance of having students develop higher order thinking skills, teacher-made tests are often heavily weighted with recall questions. Teachers who have received training in teaching and testing higher order thinking skills use fewer recall questions than those who have not received such training. Three comprehensive performance based models for assessing higher order skills have been constructed. They contain detailed procedures for constructing performance assessments in specific content domains.

Two international curricula and assessment systems are available to instruct advanced students in both subject matter content and higher order thinking skills. Students are assessed by both local school personnel and external examiners.Statewide assessment systems for elementary, middle, and high school students may contain measures of higher order thinking skills. Two states, Florida and Pennsylvania, include select, generate, and select and explain items or tasks in their mathematics, reading, and writing assessments. In Florida, higher order skills comprise at least 50% of the items or tasks, and the Pennsylvania assessments also emphasize higher order skills.A proposal for a nationwide test of higher order thinking skills and a report of a statewide pilot project for measuring such skills both emphasized the need for collaboration of subject matter experts and cognitive psychologists and/or critical thinking experts in preparing such assessments. The nationwide proposal contained a provision for multiple-choice and multiplerating items to be administered to all students and essay items to be given to representative samples of students. The pilot project contained only multiple-choice items. One study for the Florida Department of Education demonstrated the feasibility of statewide assessment of the

real-world problem-solving skills of samples of students. Instruments for implementing the proposed assessments of the skills of elementary, middle, and high school students would consist of multiple-choice, constructed-response, and possibly computeradministered performance tasks. Example items and tasks reveal a wide variety of ways that have been used to measure higher order skills. They also demonstrate the difficulties inherent in the construction of instruments to assess these skills for students at all educational levels. It should be possible to construct measures for them through the use of carefully developed item/test specifications and adequate pilot testing and analysis.

Purposes of Critical Thinking Assessment

In developing a critical thinking test we must also have a clear idea of the purpose for which the test is to be used. A variety of possible purposes exist, but no one test or assessment procedure fits them all. Here are some major possible purposes of a critical thinking test

1. Diagnosing the levels of students' critical thinking.
2. Giving students feedback about their critical thinking prowess.

3. Motivating students to be better at critical thinking.
4. Informing teachers about the success of their effforts to teach students to think critically.
5. Doing research about critical thinking instructional questions and issues.
6. Providing help in deciding whether a student should enter an eductional program.
7. Providing information for holding schools accountable for the critical thinking prowess of their students.

Why Do We Need Critical Thinkers Now?

In this knowledge age, technological search engines abound. Teachers need to cultivate students as the decision-makers who interact critically and creatively with these search engines. There are challenges we need to face. The next generation should possess the skills to be able to respond to them creatively. Now that we have access to more information worldwide than ever before, we need to teach our students to evaluate information and arguments so they can become active, intelligent, and persuasive participants in politics, science, business, the arts and society. Thus, there is a need to assess critical thinking among learners so

that instructional decisions and activities are modified accordingly.

Challenges in Assessing Critical Thinking

There are a number of challenges in assessing critical thinking skills and dispositions in students. Researchers have pointed out problems associated with both reliability and validity of existing measures. Moss and Koziol (1991) factor analyzed scores from a set of writing tasks intended to measure the critical thinking skills of students in grades 5, 8, and 11 in the context of social studies. Students who read a social studies passage either supported an inference with argumentation or evaluated an argument from the passage. The authors found no clear, common factor underlying performance across tasks that were designed to be parallel.

Further, students' abilities to use topic statements, evidence, explanations, conclusions, and logical organization did not generalize across tasks, suggesting that idiosyncratic and perhaps construct-irrelevant features of each passage or task were more salient aspects of student performance than any general ability to think critically. Silva (2008) has noted that performance-based assessments of creativity introduce, rather, subjectivity and error.

Moreover, use of such performance tasks to assess the growth of critical thinking skills over time remains fraught with difficulties as long as individual tasks communicate more noise than signal (Moss & Koziol, 1991).

Norris (1989) argues that the fact that the degree of domain specificity in critical thinking remains unresolved makes assessment of critical thinking difficult. First, the type of inferences one is trying to make remains unclear to the extent that researchers cannot agree whether critical thinking is general or subject-specific. Second, it is difficult to assess critical thinking transfer, because transfer to other contexts is confounded with subject-specific knowledge that is necessary for exercising critical thinking. Thus, a student who fails to transfer to another subject either requires additional instruction in critical thinking or additional instruction in the subject matter. Similarly, the disposition to think critically is confounded with the ability to do so. Thus, despite the fact that researchers have identified critical thinking skills and dispositions as distinct from one another, delineating their separate effects using an assessment is difficult in practice. Finally, Norris argues that traditional assessment formats are ill-suited for testing even limited aspects of

the construct. Standardized instruments using multiple-choice formats to assess credibility judgment or deductive reasoning are as likely to reflect extraneous constructs such as test-makers' empirical, religious, or political beliefs and judgments—as they are to reflect critical thinking.

Existing published assessments of critical thinking are numerous, and include the California Critical Thinking Skills Test (Facione, 1990), the Cornell Critical Thinking Tests (Ennis & Millman, 2005), the Ennis-Weir Critical Thinking Essay Test (Ennis & Weir, 1985), and the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980). As Ku (2009) points out, these instruments vary widely in both purpose and item format. However, as Kennedy et al. (1991) note, none of these tests are intended for use with students below the fourth-grade level.

Assessment Recommendations

Researchers have made several suggestions for designing assessments ideally suited to assess critical thinking skills. First, open-ended problem types may be more appropriate for assessing critical thinking than traditional multiple-choice formats. As Ku (2009) argues, available empirical evidence suggests that open-ended measures better

capture the construct of critical thinking because they are more sensitive to the dispositional aspects of critical thinking than are multiple-choice measures. For this reason, Ku recommends using tests of mixed item format, both multiple-choice and open-ended, to more completely represent both the cognitive and dispositional aspects of critical thinking. Ku (2009) argues that teachers should adopt different assessment methods, such as exercises that allow students to self-construct answers, assignments that facilitate the practice of strategic use of thinking skills in everyday contexts, and when adopting multiple-choice exercises, follow-up questions should be given to probe students' underlying reasoning.

Assessment tasks should also reflect authentic problem contexts and performances (Bonk & Smith, 1998; Halpern, 1998). This means that assessments should be based on simulations that approximate real-world problems and issues. Assessments should also use ill-structured problems, which Moss and Koziol (1991) explain to mean that test questions should require students to go beyond the available information in the task to draw inferences or make evaluations. In addition, problems should have more than one plausible or defensible solution, and there should be

sufficient information and evidence within the task to enable students to support multiple views (Moss & Koziol, 1991).

Fischer, Spiker, and Riedel (2009) argue that critical thinking is a "stimulus-bound phenomenon," meaning that certain external task features may impact whether critical thinking is elicited in a given assessment context. The authors identify a number of context variables that affect one's use of critical thinking. Stimulus characteristics focus on whether the stimuli present a set of materials that is orderly, well-organized, and coherent, or a set of materials that is uncertain, ambiguous, disorganized, and contradictory. In experimental studies that attempted to validate their model of critical thinking, Fischer et al. (2009) demonstrated that some contextual stimulus variables do seem to matter, whereas others do not. They assert that the level of substance of stimulus text—in terms of the number of unique propositions contained in that text—had no main effect on the subjects' propensity to use critical thinking, operationalized in this study as the number of questions of belief and checks on thinking observed during "think-aloud" procedures. However, the level of consistency, or lack of contradictions, within stimulus materials did have a main effect, with inconsistent or contradictory materials

more likely to prompt critical thinking than consistent and coherent stimulus materials.

Moreover, Fischer et al. (2009) demonstrated that certain types of tasks are more likely to elicit critical thinking than others. For example, tasks requiring the exercise of judgment were better for assessing critical thinking than tasks focused on simply understanding material presented in stimulus text. In particular, a task requiring examinees to either accept or reject a manuscript for publication elicited more questions of belief and checks on thinking than a task asking examinees to identify the main topic of a set of materials or to explain a scientific study described in stimulus materials.

Moss and Koziol (1991) advocate evaluating students on the basis of the quality of the arguments underlying their position, rather than the "correctness" of the answer. Lewis and Smith (1993) point out that assessment tasks must go beyond requiring simple recall of learned information. Rather, tasks should require students to manipulate what they learned in new or novel contexts. Another suggestion is that critical thinking assessments should make student reasoning visible. For example, Norris (1989) argues that testing validly for critical thinking requires that we observe an

examinee's process of thinking. One recommendation for accomplishing this in the context of a multiple-choice test is to require students to provide a rationale or justification for their choice, an idea that was repeated by Kennedy et al. (1991). Silva (2008) argued that new assessment modes are needed to measure higher-order skills, identifying several examples of recent critical thinking assessments that use novel item formats. For example, the College and Work Readiness Assessment (developed by the Council for Aid and the RAND Corporation) presents students with a 90-minute task and access to a variety of written materials on the topic, which typically represents a real-world problem. Students are then asked to make judgments and formulate a solution. River City Research Project (developed within Harvard's graduate school of education with National Science Foundation funding) is an assessment and instruction program that uses an interactive, virtual environment to present middle-school students with simulated, real-world problems that they must solve through the application of the scientific process: generating hypotheses, testing hypotheses, analyzing results, and drawing inferences and conclusions. Finally, Power Source developed by researchers at the National Center for Research on Evaluation, Standards

& Student Testing (NCRESST)—is a middle-school math assessment that combines higher-order thinking skills with mastery of basic math content in the form of narrative themes or graphic novels. Students are asked to apply math principles and to explain their reasoning.

How can pupils learn to think critically? Some educators recommend teaching thinking skills directly with widely used techniques such as Productive Thinking Program or CORT (Cognitive Research Trust). Other researchers argue that learning computer programming languages such as LOGO will improve pupil's/student's minds and teach them how to think logically. Papert (1980) said that when children learn through discovery how to give instructions to computers in LOGO, "powerful intellectual skills are developed in the process."

Adams and M.J. (1989)said that an article thinking skills curricula addresses the explosion of interest in teaching of thinking skills. It also concludes that the direct teaching of thinking offers an important complement to conventional school beginners for all students and is valuable for those with scholastic difficulties.

Perkins, Jay and Tishman (1993) proposed that another way to develop students' thinking is to providde cognitive apprenticeships in analysis, problem solving and reasoning through the regular lessons of the curriculum. They also propose that teachers can do this by creating a culture of thinking in their classrooms. This means that there is a spirit of inquisitiveness and critical thinking, a respect for reasoning and creativity, and an expectation that students will learn and understand: In such a classroom, education is seen as enculturation of a broad and complex process of acquisitive knowledge and understanding consistent with Vigotsky's Theory of Mediated Learning. This theory points that we all learn language by being a member of a particular cultural group. We also learn ways of interacting, norms of appropriate behavior and many other complicated rules and procedures through living in a culture that supports certain knowledge and values.

Evaluation

Hopkins and Stanley (1981)said that evaluation is the process of summing up the results of measurements or tests, giving them some meaning based on value judgments.

Various types of tests and evaluative instruments are used to determine the extent to which learning outcomes are achieved. Final evaluation of learning outcomes provides the learner with information regarding to the nature of the learning outcomes, his progress toward attaining these outcomes, and his difficulties in moving towards them. The teacher on the other hand, gains information regarding the appropriateness and attainability of instructional objective, and the effectiveness of his teaching methods and materials. Thus, evaluation improves both the teaching and the learning process.

Further, the results of evaluation are utilized in assigning marks and reporting pupils' achievements to parents. Evaluation reports are also important in making administrative and guidance decisions that redound to improvement of the school system and the enhancement of pupil development.

Gronlund (1981) said that from the educational standpoint, evaluation may be defined as a systematic process determining the extent to which instructional objectives are achieved by pupils. From these definitions indicate that evaluation is much more comprehensive and inclusive than

measurement. Measurement is limited to the quantitative description of an attribute, and does not include qualitative descriptions and value judgments of its worth up to those.

Gronlund (1981) said that teaching, learning and evaluation are three interdependent aspects of the educative process. These interdependence is clearly seen when the main purpose of instruction is conceived in terms of helping pupils achieve a set of learning outcomes which include changes in the intellectual, emotional or physical domains. Instructional objectives or in other words, desired changes in the pupils, are brought about by planned learning activities, and pupils' progress is evaluated by tests and other devices.

Gronlund (1981)said that testing is a technique of obtaining information needed for evaluation purposes. Tests are devices used to obtained such information. Tests provide teachers with information that can aid them in improving instruction.

For school administrators, tests provide informations for planning and evaluating the effectiveness of educational programs. In general test serve three functions: improvement of instruction, administrative uses and guidance uses.

Tests can be classified in terms of their function content item forms, administrative procedures, scoring system, and interpretation of results.

Educational test is the measure of results or effects of instruction while Psychological test is the measure of intangible aspects of behavior – attitude, interest, emotional adjustment, intelligence and ability.

Characteristics of good tests are validity which refers to the extent to which the test serves its purpose or the efficiency with which it measures what it intends to measure, reliabilty, administrability, scorability, interpretability and economy.

To be effective, evaluation should be viewed as a process of determining the degree to which pupils have achieved instructional objectives. This process is most effective when based on sound principles that provide direction to the process and which can also serve as criteria for assessing procedures and practices.

Gronlund (1981) said that some principles of evaluation are: Evaluation based on clearly stated objectives, evaluation procedures and techniques selected in terms of the

clearly stated objectives or the purpose they will serve, comprehensive evaluation, continuous evaluation and as an integral part of the teaching and learning process, diagnostic and functional evaluation, cooperative evaluation and evaluation used judiciously.

Teachers have always been concerned with measuring and evaluating the progress of their students. As the goal of education have become more complex and as the number of students has increased enormously, this task has become much more difficult.

Teachers have an obligation to provide their students with the best evaluation. This implies that they must have some procedures where by they can reliably and validly evaluate how effectively their students have been taught. The classroom achievement test is one such tool.

Kelly (1965) said that asserted that the effectiveness of the teaching-learning process is evaluation. By means of evaluation, the progress of the learner’s attainment of his educational goals is ascertained and the future direction toward these educational goals is determined. Moreover, evaluation is always qualitative and in terms of educational maturity of the learner. The purpose of the evaluation are

threefold, namely: the adaptation of instruction in accordance with the differing and individual needs of the students, the educational and personal guidance of the learner, and the appraisal of the total school program. He also stressed that the process of evaluation involves four distinct operations, namely: 1) the identification of the educational goals, attributes, objectives or purposes judged to be of value worth in the specimen to be evaluated; 2) the development and employment of means to assess the educational goals or attributes validly and reliably; 3) the synthesis of the evidence yielded by these procedures into a final judgment of attainability and goodness of the derived change in behavior , and 4) the redefining of these educational goals and purposes in terms of new and desired changes or directions necessitated by these practical judgments of evidence.

Furst (1968) emphasized that evaluation can serve a variety of purposes, some central and others only perepheral or incidental. These include 1) determining the effectiveness of courses and programs, 2)testing assumptions about given practices, 3) clarifying, selecting and appraising objectives, 4)discovering and understanding learning

difficulties, 5) planning instruction, 6) motivating learning, 7) guidance, 8) selection of pupils, 9) placement, 10) certification and in – service development of teachers.

Tyler (1959) said that evaluation is a basic task of the educator. It is one of the four basic tasks in the development of curriculum or a plan of instruction. These four are: 1) to determine the objectives which course or program should seek to attain, 2) to select learning experiences which will help to bring about the attainment of these objectives, 3) to organize these learning experiences so as to provide continuity and sequence for the student and to help him integrate what otherwise appear as isolated experiences, and 4) to determining the extent to which the objectives are being attained. These view results in the assumption that education is a process for changing the behavior patterns of human beings. As a result of this process, it is expected that individuals will acquire ideas that they did not have before, improve their ways of thinking, develop tastes sensitives, modiffy their attitudes, and improve in other desired ways.

Francisco H. Cangas (1981) presented a broader view of evaluation and its place in education that is its usefulness

in the improvement of teaching and learning. These view encompasses the following concepts:

Evaluation as a method of acquiring and processing the evidence needed to improve the students' learning and the teaching.

Evaluation as a mechanism including a great variety of evidence beyond the usual final paper and pencil examination.

Evaluation as an aid in clarifying the significant goals and objectives of education and as a process for determining the extent to which students are developing in these desired ways.

Evaluation as a system of quality control in which it may be determined at each step in the teaching – learning process whether the process is effective or not, and if not, what changes must be made to ensure its effectiveness before it is too late.

Finally, evaluation as a tool in educational practice for determining whether alternative procedures are equally effective or not in achieving a set of educational ends.

Test Construction

Constructing a satisfactory test is a deliberate process and considered one of the most challenging tasks of a teacher. It demands an understanding of the objectives and content being assessed, the reading and vocabulary level of the examinees, and test – taking factors such as responses styles and test sophistication.

Hopkins and Stanley (1958) stated that a good test must be planned and careful planning must precede its construction. One must consider the topics and objectives to be measuered, the purpose the scores are to serve, and the conditions under which testing will occur. The two authors presented the following guidelines that should be followed in developing educational achievement tests:

The test should make provisions for evaluating the important objectives of instruction that are immediately measurable.

The test should reflect the content and process objectives in proportion to their importance and emphasis in the course.

The nature of the test should reflect its purpose, that is, either to assess individual differences or to certify mastery.

The test should be appropriate in length and readability level.

In the planning test, the test maker must take into consideration the objectives of administering the test. These objectives in the construction of the test must also be based on the objectives of the different lessons that were taught in the classroom. These objective can only be realised or assessed after a formal education has been concluded. General educational objectives must be specified for according to Clark from “parent objective” many “offspring” objectives can be derived at various levels of specificity. Objectives must be started in behavioral terms so that they can be attainable and measurable. In formulating objectives, one must remember the two important domains to develop, namely: cognitive and affective.

Krathwol (1964)said that the cognitive domain includes those objectives that deal with the recall of recognition of learned material and the development of intellectual abilities and skills. The largest proportion of educational

objectives fall on cognitive domain. According to him and his companions, this domain is the core of the most curriculum and test development.

He stated further that the clearest definitions of objectives for the cognitive domain are phrased as descriptions of desired student behavior, that is, in terms of knowledge, understanding and abilities that can be demonstrated.

Bloom, et. al (1965)said that the affective domain includes objectives that emphasize interests, attitudes, and values, and the development of appreciations and adequate adjustment. Objectives in this domain are not stated very precisely; and in fact, teachers do not appear to be very clear about the learning experiences which are appropriate to these objective.

After listening, analyzing, checking and defining each objective in terms of expected student outcomes, it is an inventory step to list the elements that area part of each objective. Define each element in terms of student behavior and list items of subject matter related to each element.

To avoid falling into the pitfall of constructing test based merely on factual knowledge, an evaluative criteria on skills to be developed in the students must be spelled out. Criteria, as used for evaluation is so important that a good test cannot be constructed without it.

Mangubat (1982) said that to help the test maker decide on how much emphasis be given to each objective, it is advisable to establish a table of specifications.

Terrel (1971), Classroom objectives derived from cognitive levels of learning as infered from selected teacher – made tests. The findings are:

1) There is a need to train teachers in the use of taxonomic distinctions in the test items they construct.

2) Teacher education institutions can assist in the teacher education curriculum by including objectives that require prospective teachers to demonstrate an ability to frame questions that utilize all levels of cognition.

3) Programmed materials are needed to train teachers on an individual basis. This will permit flexibility in terms of time for the improvement of test items at each level of cognition.

4) Educational institutions should develop progrms based on performance criteria. Instructional objectives should be specifically stated. Through such planning, instructional objectives and teacher-made tests are likely to reflect all levels of cognition.

Planning the Test

Good tests do not just happen . They require adequate and extensive planning so that the goals of instruction, the teaching strategy to be employed, the textual materials, and the evaluative procedure are all related in some meaningful fashion. Most teachers recognize the importance of having some systematic procedure for ascertaining the extent to which their instructional objectives have been realized by their pupils.

To ensure successful planning, the teacher can follow these steps in order to come up with a good test a) determining the objectives, b) prepairing the Table of Specifications, c) selecting the appropriate item format, d) writing the test items, and e) editing the test items

Test Format

After planning the content and cognitive objectives for the test, decisions must be made the best way or format to measure them. The format refers to whether the test will be objective, i.e., multiple choice, true false, matching, etc., or essay type. There are factors to consider when deciding on the format of a test. The format should be most appropriate for measuring the cognitive objectives of the test.

Requirements of a Good Test

A test must possess certain requirements before use. Gronlund and Linn (2000) asserted that a good test must possess validity, reliabilty, objectivity, discrimination and ease of administration, and scoring. All these factors should be present in a good test because they are interdependent and they affect each other.

Validity is the most important feature of a good test. A test is valid when it measures well what it intends to measure. A single item is valid when it does the job expected of it. Kelly (1965) said that a good test, as well as good teaching, should penalize rote learning, rather than put a premium upon it. It should be the teacher's objective in

tests construction to phrase or present the questions and responses that only a genuine understanding of the concepts will enable the student to respond correctly.

The approach to validity from the standpoint of adjustment to life and success to school achievement is illustrated in L.M. Terman's studies of gifted children which showed that the concept of validity is illustrated by the fact that intelligence tests are considered to be valid when their results corresponds to scores on other recognized intelligence tests, to pooled judgment of a panel of experts who know the examinees, to an estimate of the subject's intelligence as inferred from measures of his adjustment in various situations or to the subject's achievement in school.

According to Hopkins and Stanley (1996), the validity of a measure is how well it fulfills the function for which it is being used. Regardless of the other merits of a test, if it lacks validity, the information it provides is useless. The validity of a test can be viewed as the accuracy of specified inferrences made from it's scores. These inferences pertain to (1) performance on a "universe" of items (content validity), (2) performance on some criterion (criterion related validity), (3) the degree to which certain

psychological traits or constructs are actually represented by test performance (construct validity). During the process of test validation one examines the relationships between the test scores and other empirical data and other logical considerations. The extent to which a test improves the accuracy of decisions is the extent to which it provides useful information and, hence, has practical validity. If a test possesses many validities, it may be highly valid for one purpose but not for others. The question of test validity is always particularized, it can be answered only in relation to a given specific task for a given population of examiners.

The same authors emphasized that there are factors that prevent the test from measuring what it is supposed to measure. These factors are 1) poor reading ability, 2) limited vocabulary, 3) difficulty in memorizing, and 4) slow reaction. These and other similar factors tend to penalize certain students in many tests. The total scores of such students will not be a valid indication of their real achievement.

Gronlund and Linn (2000) said that a reliable test is trustworthy, accurate and consistent. If the test measures in exactly the same manner each time it is administered and if

the factors that affect the test scores affect them in the same extent everytime the test is given, then the test is said to be very high in reliability. In other words, a highly reliable test should yield the same score when administered twice to the same student provided of course, that no learning occurs while the test is being taken the first time or no learning or forgetting takes place between testings.

According to Baron and Bernard (1958) objectivity is another requirement of a test. Objectivity reduces the subjective element in judgment to a minimum. The teacher achieves objectivity by using the scoring key provided with the test by careful adhering to printed directions for administration, scoring and interpretation.

Discrimination should also be possessed by an achievement test in order to be considered a good one. It refers to the ability of a total test or individual item to differentiate effectively between the good and poor students.

Gronlund and Linn (2000) added that in constructing an achievement test, it is equally important to sample liberally all phases of instruction which are supposed to be covered. In other words, the test must be comprehensive. The test

should be considered good, likewise, must be easy to administer and easy to score.

Making Your Own Test

In making a test, it is probably better that it be at least somewhat open ended. It has advantages as opposed to the multiple-choice tests that is difficult and time consuming to develop as it requires a series of revisions, tryouts, and more revisions. Open-ended assessment is better adaptes to do-it-yourself test makers and can be more comprehensive. Norris and Ennis (1989) make suggestions for open-ended assessment.

Effects of Evaluation

Norman (1967) asserted that in education, the process of studying or testing may have so much effect on students, teachers and others that what one is investigating cannot be completely separated from the investigation process. To measure a child's intelligence is to applicably change the child and his parents, as well as to affect the way in which teachers and others come to view the child. Thus, the process of evaluation has an effect on the learner and the teacher, as well as on others involved in education. Goslin said that

it is necessary to understand the possible effects and to deal with them intelligently. If these effects are understood and utilized properly, they can do much to enhance student's learning as well as his regard for himself. If the effects are not used well, they can do great damage to the student as well as to the educational system. The point to be made is that the effects of evaluation can be maximized or minimized, but they cannot be entirely controlled. Also the effects can be positive or negative, and only rarely can they be entirely neutralized.

Bloom (1969) pointed out the following ideas about the effects of evaluation:

1) Examinations which are regarded as measuring important and relatively stable characteristics of the individual have the greatest effect. Thus, the I.Q. score is regarded as important because it is widely believed that it is highly stable and what little can be done to alter it. While each of these assumptions can, in part, be questionned, school people and laymen believe that the intelligence test so important. Other tests which may be regarded in the same way are aptitude tests, personality measures, vocational interest tests, and sometimes, tests of a reading ability.

2) Examinations which are used to make important decisions at major disjunctions in the educational system also have great effects. Examinations are used for certification of the completion of an educational program, for the selection of students for particular programs or streams of education, and for the determination of admission to advanced programs of education. There are critical in the lives and careers of students, the educational system, and the entire society. Such examinations have a profound effect on the curriculum and instruction. They determine who will have certain opportunities and who will not, and they determine the type of person who will be admitted toparticular occupations and professions. Examinations help to shape the society's view of itself and of the variations among the members of the society.

3) Examinations for which the results become part of the student's permanent record or which are made public also have great effects. If the results of a particular examination are reffered to repeatedly in making critical decisions about the student, the examination will have a marked influence on the student and on the adults who are interested in or concerned with him. Here, the whole issue is invasion of privacy. In so

far as examination results are made a part of the permanent record of the student and such a record may be available to other teachers and school authorities, they must be regarded as important by the student, and to do well or poorly on them is a matter of vital concern to the student and others.

4) Examinations which are used to judge the effectiveness of teachers, schools or systems of education have great effect on the institution involved. While such examinations may have little but direct influence on the students tested, they may have marked influence on the curriculum and on the staff and administration of the institution. Undoubtedly, the responsiveness of the institution and staff to the examination results is determined by the extent to which the results can be related to specific practices and persons in the institutions.

5) Perhaps the least effect is likely when the examination results are not related to individuals, practices or institutions. If the examinations are related to anonymous individuals and institutions with a minimum of comparison among individuals or institutions with a minimum of publicity, the results are unlikely to have much effect on either individuals or institutions.

6) Examinations are likely to have little effect if they are considered to be measuring trivial things which are not regarded as important by the students, teachers and others. Thus, a test of handwriting elegance is unlikely to have much influence on students or schools at present, while sixty years ago it might have been regardedas of the utmost importance.

7) Finally, examinations which are not used for making significant decisions by or about individuals or institutions, and where this is known in advance by the examinees, are likely to have minimal effects on the individuals and institutions.

Perhaps, the main point to be made about the effects of examinations is that it is largely a perceptual phenomenon. That is ,if students, teachers ,or administrators believe that the results of an examination are important, it matters very little whether this is really true or false – the effect is produced by what individuals perceive to be the case.

CHAPTER III

METHODOLOGY

This chapter presents the method and procedures employed in the collection, analysis and interpretation of data required in the study. The discussions are sequenced as follows; Research Design, Participants of the study , Research Instrument , Data Gathering Procedure and Data Analyis

Research Design

This study made use of the descriptive method of research. The descriptive method was adopted in the study since it seeks to know the conditions typically existing at the time of the study. It involve the construction of a critical thinking test that employed the standard procedure in test construction, estblishment of content and construct validity and establishment of test reliability. This study aimed to developed and evaluate a critical thinking test in Grade III Mathematics.

Participants of the Study

The pupil – respondents in this study were grade three pupils of different sections. Table 1 shows these sections and the number of respondents from each of them.

Table 1
Frequency Distribution of Respondents
SY. 2012 – 2013

Name of Section	**Population**		
	Boys	Girls	Total
Fast Learner	20	20	40
Section 2	20	20	40
Total	40	40	80

Of the 105 pupils from the two sections, 80 pupils were selected by random sampling.

Permission for the tryout in each section was obtained from the proper school authorities. A letter requesting the permission of the principal was made for the administration of (Appendix B.1).

Research Instrument

The developed instrument in this study consisted of twenty-five item critical thinking test in selected topics in Mathematics for Grade three. The test seeks to measure pupils in terms of: a)defining and clarifying the problem

b)judging information related to the problem and c) solving problems/drawing conclusions. The Items are categorized under the Table 9.7 of Critical Thinking Skills (Kneedler,1985).

The developed test was validated by one professor of the Philippine Normal University (PNU) and a Division Supervisor of Mathematics in the Division of Santa Rosa Laguna. Based on the reccomendation of the validators, the test was modified. The final form of tha test was evaluated by four eperts with masters degree in psychology.

Test Development Procedure

Preparing the Table of Specification

The Table of Specification was based on the number of times each critical thinking skill in Kneedler's criteria a) defining and clarifying the problem, b) judging information related to the problem, and c) solving problem/drawing conclusion was used in the problem (test). The Table of Specifications shows the distribution of the skills test items, the item placement and the critical thinking skills category where the item falls.

Table of Specifications for the Critical Thinking Skill Test (CTST)

Critical Thinking Skills (Table 9.7)	No. of Times each Critical Thinking is used in the problem (Test)		
Content	Frequency	%	Item number
A. Defining and Clarifying the Problem			
1. identify central issues or problems	25	25	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25
2. Compare similarities and differences	15	15.15	2,3,4,5,6,7,16,17,18,19,20,22,23,24,25
3. Determine which information is relevant	23	23.23	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,20,22,23,24,25
4. Formulate Appropriate questions	9	.9	3,7,16,19,21,22,23,24,25
B. Judging Information Related to the Problem			
5. Check Consistency	1	1	18
C. Solving problems/ drawing Conclusions			
6. Recognize the adequacy of data	22	22.22	3,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25
7. Predict probable consequences	4	4	22,23,24,25
TOTAL	99	100%	

Constructing the Test

In developing the test the following steps were followed.

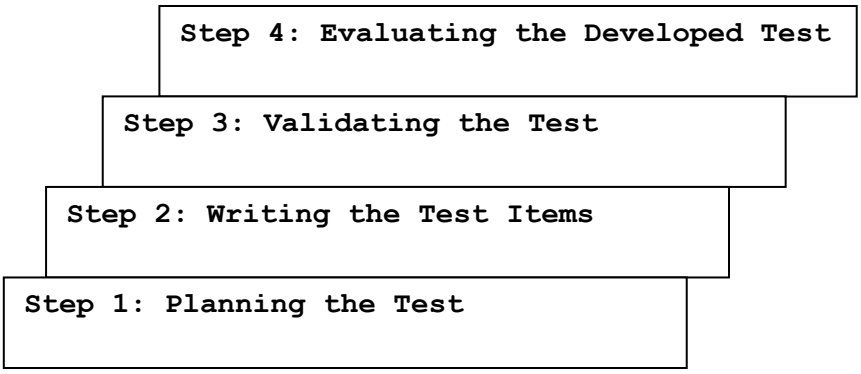


Figure 2

1. Planning the Test

Books, journals, and research outputs related to the study were read and asked for copies of textbooks that are currently used by pupils with the help of elementary teachers. These materials serve as guides for the development of the test. Then the test format was selected.

2. Test Writing Stage

A table of specification was prepared to allocate items that represent Kneedler's criteria. The researcher designed a problem solving test. The critical thinking

test was consisting of selected topics in Mathematics for Grade III.

3. Validating the Test

The preliminary form of the test was shown to the paper adviser for her comments and suggestions. The test was then administered to a group of Grade III pupils in a well-ventilated room under properly lighted conditions. The pupils were observed carefully to determine what problem they may encounter while taking the test. The papers were corrected and subjected to items analysis. Based on the item analysis and observations, the test was revised.

The revised test was shown to two holders of Masters Degree in Psychology for item validation, among their comments were:

1. Children not familiar with the terms
2. Wrong grammar
3. Items not simple to read
4. Proper use of number word

Those items falling under any of the above categories were either recast or entirely changed as suggested by the adviser. These are the following

Original Items	**New Items**
3. If twice Nena's age is increased by 8, the result is 26. How old is Nena?	3. If twice Nena's age increased by 8 is 26. How old is Nena?
4. A cooking class baked 95 cookies. They served 2/5 of the cookies to visitors. How many cookies remained?	4. A cooking class baked 95 cookies. They served 2/5 of the cookies to visitors. How many cookies were left?
9. Because of lack of books, 12 books are shared by 20 pupils. What is the ratio, in lowest terms, of books to pupils?	9. Because of lack of books, twelve books are shared by 20 pupils. What is the ratio of books to pupils in lowest terms?
15. What 2 numbers have a sum of 21 and product of 108?	15. What two numbers have a sum of 21 and a product of 108?
25. What is the biggest number you can write using 0,5,8,7?	25. What is the biggest number you can form using 0,5,8,7?

The item validators were asked to specify their level of agreement or observation to each of a list of items. The scale is appropriate or not appropriate (Appendix D).

4. Evaluating the Developed Test

The revised test was submitted to four subject experts, two from the Math Department of PNU, an Educator Program Supervisor in Mathematics of the Division of Santa Rosa City, Laguna a principal and directress of Laguna Eastern Academy Santa Rosa City Laguna for evaluation using an appropriate form (Appendix C). The expert used the Likert scale to evaluate the developed test the scale is as follows:

	Scale Value
Strongly Agree	4
Agree	3
Disagree	2
Strongly Agree	1

After the test items had been evaluated and make the changes, the test was administered to the final respondents of this study, composed of 80 Grade – III pupils of Santa Rosa Elementary School S.Y. 2012-2013.

Guide in the Administration and Scoring of the Test.

1. This test is good for one hour only.

2. In scoring the test, the items measuring a specific skill are found in the following

Table 3
Table of Specification of the Revised Critical Thinking Skills Test

A. Defining and Clarifying the Problem	**Item number**
1. identify central issues or problems	1,2,3,4,5,6,7,8,9, 10,11,12,13,14,15, 16,17,18,19,20,21, 22,23,24,25
2. Compare similarities and differences	2,3,4,5,6,7,16,17, 18,19,20, 22,23,24,25
3. Determine which information is relevant	1,2,3,4,5,6,7,8,9, 10,11,12,13,14,15, 16,17,18,20,22,23, 24,25
4. Formulate Appropriate questions	3,7,16,19,21,22,23 ,24,25
B. Judging Information Related to the Problem	
5. Check Consistency	18
C. Solving problems/ drawing Conclusions	
6. Recognize the adequacy of data	3,5,6,7,8,9,10,11, 12,13,14,15,16,17, 18,19,20,21,22,23, 24,25
7. Predict probable consequences	22,23,24,25

Finally, collect, connect and encoded the results.

CHAPTER IV

PRESENTATION, ANALYSIS AND INTERPRETATION OF DATA

This chapter presents the analysis and interpretation of the data based on the sequence of questions in the statement of the purpose.

Purpose 1: Construct Test Items that will measure Critical Thinking Skills in Mathematics among Grade-III pupils in terms of .

1.1 Defining and Clarifying the Problem

1.2 Judging Information Related to the Problem

1.3 Solving Problems/Drawing Conclusions

In constructing a test items that will measure critical thinking skill in mathematics among grade three pupils the following steps were undertaken:

Step 1 : Read Books, Journals, and Research Output (Thesis) choose the purpose of the instrument.

Step 2 : Review, Revise, and Tryout. Identify the construct(s) and construct initial item pool.

Step 3 : Validation of the Test.

Step 4 : Evaluation of the Developed Test.

Purpose 2: Evaluate the developed test in terms of content

Table 4

Evaluation of the experts on the developed critical thinking skills test in terms of content.

Experts' Evaluation	Average Means	Interpretation
Evaluator 1	3.73	Highly Agree
Evaluator 2	3.87	Highly Agree
Evaluator 3	3.80	Highly Agree
Evaluator 4	3.80	Highly Agree
Overall Average	**3.80**	**Highly Agree**

Scale

3.25 – 4.00 – Highly Agree

2.50 – 3.24 - Agree

1.75 – 2.49 - Disagree

1.00 – 1.74 – Highly disagree

It can be noted from the table that the overall average of 3.80, the evaluators *highly agree* that the test truly measures the critical thinking skills of grade III learners in Mathematics.

Purpose 3: Estimate the reliability of the test using test-retest method.

Table 5

Test scores of 80 pupils on the first tryout and final tryout of the test.

Pupils	1st Trial	2nd Trial		Pupils	1st Trial	2nd Trial
1	18	24		**41**	17	20
2	17	17		**42**	14	17
3	17	16		**43**	18	22
4	19	22		**44**	17	17
5	13	18		**45**	16	19
6	19	19		**46**	19	21
7	14	17		**47**	16	18
8	17	19		**48**	17	20
9	18	21		**49**	19	19
10	11	18		**50**	14	18
11	21	23		**51**	16	18
12	14	16		**52**	21	23
13	18	19		**53**	18	21
14	13	19		**54**	17	20
15	21	21		**55**	14	18
16	19	18		**56**	18	22
17	20	21		**57**	16	20
18	15	19		**58**	15	18
19	20	20		**59**	20	21
20	17	19		**60**	17	19
21	22	23		**61**	18	21
22	18	17		**62**	19	20
23	20	23		**63**	14	18
24	20	21		**64**	17	20
25	19	21		**65**	16	20
26	15	19		**66**	18	22
27	19	19		**67**	18	20
28	15	17		**68**	14	17
29	21	23		**69**	17	20
30	14	18		**70**	14	17
31	18	18		**71**	19	23
32	18	22		**72**	12	14
33	17	20		**73**	20	23
34	18	19		**74**	17	21
35	13	18		**75**	16	15
36	15	17		**76**	22	23
37	18	17		**77**	10	16
38	20	23		**78**	20	20
39	18	21		**79**	19	22
40	20	21		**80**	20	20

As reflected in the table, the first column represents the number of pupils used in the administration of the first and second try-out of the test. Under the second and the third column were the scores of 80 pupils in the first and second try-out of the test.

Table 6
Means and Standard Deviation of the
(First and Second Try – out)

Tryout	N	Minimum	Maximum	Mean	Std. Deviation
First	**80**	**10.00**	**22.00**	**17.2250**	**2.56547**
Second	**80**	**14.00**	**24.00**	**19.5750**	**2.18612**
Valid N (listwise)	**80**				

Table 7
Pearson Correlations (First and Second Try – out)

		First	Second
First	Pearson correlation	1	.728
	Sig. (2-tailed)		.000
	N	80	80
Second	Pearson correlation	.728	1
	Sig. (2-tailed)	.000	
	N	80	80

By using the Pearson-Product Moment Coefficient of Correlation, table 6 and 7 shows the correlation between the first tryout and the final tryout of the test was

computed to be .73, which means that the test was highly correlated and was very reliable.

The criteria for the interpretation of the Pearson r in terms of verbal description were used as follows:

.80 – 1.00 - Very High Correlation

.60 - .79 - High Correlation

.40 - .59 - Modest Correlation

.20 - .39 - Slight Correlation

.01 - .19 - Very Slight Correlation

The final form of the critical thinking test in Mathematics for Grade III consists of twenty-five items (Appendix E).

CHAPTER V

Summary, Results and Recommendations

This chapter presents the summary of findings, the conclusions based on findings and the recommendations based on conclusions.

Summary

This study was conducted for the purpose of developing and evaluating a critical thinking test for grade III pupils.

Specifically, this study sought to:

1.Construct Test Items that will measure critical thinking skills in Mathematics among grade III pupils in terms of 1.1 defining and clarifying the problem, 1.2 judging information related to problem, and 1.3 solving problems/drawing conclusions

A systematic procedure was followed in developing and evaluating the critical thinking test on Mathematics. The study followed four steps of test development. 1) Books, journals, and research outputs related to the study were read and asked for copies of textbooks that are currently used by pupils with the help of elementary teachers. These materials serve as guides for the development of the test.

Stating the purpose of the instrument, and making table of specification. Then the test format was selected. 2) Test instrument writing stage which include identifying the construct, developing items based on the purpose of the study, based on who will take the test and based on the steps of proper test development. 3) Review, revise and Try-outs. The researcher administered the first try-out in Santa Rosa Elementary School Central - I. After the test, the adviser and test experts reviewed and revised items. From the first try-out five items were revised and twenty-items were retained (See table 3). 4) establishing the test validity and reliability.

2. Evaluate the developed test in terms of content.

   The researcher consulted four experts to rate the critical thinking test in terms of the fifteen items critical thinking content. Evaluator one (3.73), evaluator two (3.87), Evaluator three (3.80) and evaluator four (3.80). The average of all weighted means was 3.80 which implies that the test is excellent.

3. Estimate the reliability of the test using test- re-test method.

To establish the reliabilty of the test the Pearson Product-Moment Correlation was used. The computed Reliability Coefficient was .73 which means that the test was very reliable.

Findings

The following were the findings in the study:

1. The evaluators *strongly agree* that the developed test items truly measures the critical thinking skills of Grade III pupils.
2. The developed test has a high reliability index of .73. Since the test scores were correlated, this is another indicator of the validity of the test.

Conclusions

In light of the findings of the study, these conclusions were made:

1. The developed critical thinking skills test is valid and reliable.
2. The developed test is usable and may be used to measure the critical thinking skills of Grade III pupils.

Recommendations

In light of the findings and conclusions of the study, the following recommendations are given

1. Subject the developed instrument to further validity and reliability.
2. The instrument can be used as a critical thinking test for Grade III pupils to approximate their performance in Mathematics.
3. To develop other critical thinking tests appropriate for other grade levels.

B I B L I O G R A P H Y

American Educational Research Association (AERA), Psychological Association (PA) and National Council on Measurement Education (NCME), ***Standards for Educational and Psychological Testing***, Washington, D.C.: AERA, 1999. Retrieved July 27,2012 from www.apa.org>Science.Directorate>Programs.

Askvill-Amazon.com,(2012). ***How can critical thinking be applied in Math***. Retrieved July 19,2012 from http://askville.amazon.com/critical-thinking-give-description/AnswerViewer.do?requestId=51589486

Bailin, S. (2002). ***Critical thinking and science education***. Science & Education. Journal of Curriculum Studies.

Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). ***Conceptualizing critical thinking***. Journal of Curriculum Studies.

Baron, Dennis and Harold W. Bernard. ***Evaluating Techniques For Classroom Teachers***. New York: McGraw Hill Book Co., 1978

Bloom, Benjamin S., ***Taxonomy of Educational Objectives*** Published by Allyn and Bacon, MA Copyright (c) 1984 by Pearson Education.

Bonk, C. J., & Smith, G. S. (1998). ***Alternative instructional strategies for creative and critical thinking in the accounting curriculum.*** Journal of Accounting Education.

Braun, H. and Wainer, H. (1988). ***Test validity***. Hillsdale, NJ : Lawrence Erlbaum Associates

Brown, A. L. (1990). ***Domain specific principles affect learning and transfer in children***. Cognitive Science.1990.

Cangas, Francisco H. ***"The New Concept of Evaluation of Pupil Learning,"*** The Philippine Journal of Education. LX, July 1981.

Case, R. (2005). ***Moving critical thinking to the main stage.*** Education Canada.

Casta, A. (Ed.) ***Developing Minds: A Resource Book for Teaching Thinking***, California: by the Association for Supervision and Curriculum Development, 1985

Center for Development of Teaching and Learning.(2003).***Critical Thinking and Pedagogy.*** Retrieved September 12,2012 from http://www.cdtl.nusedu.sg/ctp/maths.htm

Clark, D. C. ***Using Instructional Objectives in Teaching.*** Glenview, Illinois: Scot Foreman,1972.

Cross, D. R., & Paris, S. G. (1988). ***Developmental and instructional analysis of children's metacognition and reading comprehension***. Journal of Educational Psychology.

Dewey, J. (1910). ***How we think***. Boston: D.C.Health. Educational Policies Commission. (1961). ***The Central purpose of American Education.*** Washington, DC: National Education Association.

Dick, Jeremy and Faivre Alain, ***Automating the generation and sequencing of test cases from model-based specifications***. Springer Berlin Heidelburg Publishing, 1993

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). ***The evolution of research on collaborative learning***. In E. Spada & P. Reiman (Eds.), ***Learning in humans and machine: Towards an interdisciplinary learning science*)**. Oxford, England: Elsevier.

Dressel. Paul L. ***"Information Which Should be Provided By Test Publishers and Testing Agencies on the Validity and Use of their Tests, Achievement Tests."*** Proceedings. Invitational

Conference on Testing Problems. Princetown: Educational Testing Service, 1970

Downie, n. and Health, R. (1983). ***Basic statistical methods***. New York: Harper and Harper and Row, Publisher, Inc.

Ennis, R. H. (1985). ***A logical basis for measuring critical thinking skills***. Educational Leadership.

Ennis, R. H. (1989). ***Critical thinking and subject specificity***: Clarification and needed research. Educational Researcher.

Ennis, R. H. (1993). **Critical thinking assessment**. Theory into Practice. Retrieved August 7,2012 from http://www.3.qcc.cuny.edu/wikifiles/file/Ennis%20Critical%20Thinking%20Assessment.pdf

Enright, B. E., & Beattie, S. A. (1992). **Assessing critical thinking in mathematics**. Diagnostique 17(2), 137-144.

Facione, P. A. (1990). ***Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction***. Millbrae, CA: The California Academic Press.

Facione, P. A. (2011). ***What It Is and Why It Counts. Englewood Cliffs.N.U.*** Retrieved August 27,2012 from www.slideshare.net/adindasafrini/criticalthinkingwhati t-is-and-why-it-counts

Facione, P. A. (2000). ***The disposition toward critical thinking: Its character, measurement, and relation to critical thinking skill***. Informal Logic.

FJ King; Goodson L. Rohani.F. ***Higher Order Thinking Skills***, The Educational Sources Program. Retrieved July 12,2012. Retrieved from www. Cala.fsu.edu.

Foxcroft C., Paterson, H., Le Reux N. And Herbst, D., ***Human Sciences:*** Anastaci and Urbina, 1997

Fraenkel,J. and Wallen, N. 2008. ***How to design and evaluate research in education*** . New York: McGraw Hill Companies, Inc.

Furst. Edward J. ***Constructing Evaluation Instrument***. New York: Longsmans, Green and Company 1988.

Gellin, A. (2003). ***The effect of undergraduate student involvement on critical thinking***: A meta-analysis of the literature 1991–2000. Journal of College Student Development.

Gelman, S. A., & Markman, E. M. (1986). ***Categories and induction in young children***. Cognition.

Gronlund, N. and Linn, R. (2000). ***Measurement and assessment in teaching* (8th Ed.)**. Upper saddle River, New Jersey; Prentice-Hall, Inc. Pearson Education.

Gronlund, Norman Edward, ***Measurement and Evaluation in Teaching***, New York : Mc Millan, 1967

Gronlund, Norman Edward, ***Measurement and Evaluation in Teaching***, New York : Mc Millan, 1981

Halpern, D. F. (1998). ***Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring***. American Psychologist.

Halpern, D. F. (2001) ***Assessing the effectiveness of critical thinking instruction***. The Journal of General Education.

Hopkin, K. D., and Stanley, J.C. (1990) ***Educational and Psychological Measurement and Evaluation***. Prentice Hall, 1990. Retrieved November 17, 2012 from www.google.com.ph/books/about/Education_and_psychologicalandmeasurement.htm

Jaswal, V. K., & Neely, L. A. (2006). ***Adults don't always know best: Preschoolers use past reliability over age when learning new words.*** Psychological Science.

John D., and Lexington. Mass;Health, 1982. **How We Think, *Importance in Schooling.*** Gerund Principle in Achieving Critical Thinking.

Kansas State University. ***Table of Specification.*** Retrieved October 16,2012 from http://www.specialconnections.ku.edu/cgi-bin/cgiwrap/speconn/main.php?cat=assessment7&section=main&subsection=qualitytest/table

Kelly, William A. ***Educational Psychology***, New York: The Bouce Publishing Co., 1965

Kneedler, P. ***California Assesses Critical Thinking.*** California: A Pearson Publishing Co., 1985

Kennedy, M., Fisher, M. B., & Ennis, R. H. (1991). ***Critical thinking: Literature review and needed research.*** In L. Idol & B.F. Jones (Eds.), Educational values and cognitive instruction: Implications for reform. Hillsdale, New Jersey: Lawrence Erlbaum & Associates.

Koenig, M. A., & Harris, P. L. (2005). ***Preschoolers mistrust ignorant and inaccurate speakers.*** Child Development.

Krathwohl. D. R.et. al. ***Taxonomy of Educational Objectives: Handbook II, Cognitive Domain.*** New York: David McKay 1964

Lewis, A., & Smith, D. (1993). ***Defining higher order thinking.*** Theory into Practice.

Mangubat, Emmanuel. ***"How to Construct Tests Based on a Table of Specifications,"*** The Modern Teacher, XXVI, August 1982.

Marsh, H.W. ***Self-Concept: Its Multifaceted Hierarchial Structure.*** California: Copyright by Laurence Erbaum Associates, Inc., 1985

Marshall H. ***Reconceptualizing Learning for Restructured Schools.*** California: April 1992

Max Horkheimer (1937). ***Traditional and Critical Theory.****California.*

McPeck, J. E. (1990). ***Critical thinking and subject specificity:*** *A reply to Ennis*. Educational Researcher.

Moore, Brooke Noel; Parker Richard. ***Critical Thinking***, Eighth Edition, California State University, Ohio; Mc Graw – Hitt Companies Inc., 2007

Norris, S. P. (1989). ***Can we test validly for critical thinking?*** Educational Researcher.

Office of Educational Assessment University of Washington. ***Item Analysis***. Retrieved August 6,2012 from http://www.washington.edu/oea/pdfs/resources/item analysis .pdf

Oriondo, L.L. Dallo-Antonio E.M. ***Evaluating Educational Outcome.*** Published and Copyright, by Rex Book store, 1984

Papert, S. (1980). ***A Critical Thinking Article Issue of the Chronicle of Higher Education***: April 28,1995

Pearson's Research Reports, ***Critical Thinking: A Literature Review*** .Retrieved August17,2012 from http://www.pearsonassessments.com/research.

Perkins, D. N., Allen, R., & Hafner, J. (1983). ***Difficulties in everyday reasoning.*** In W. Maxwell (Ed.), *Thinking*: The frontier expands. Hillsdale, New Jersey: Lawrence Erlbaum & Associates.
Silva, E. (2008). *Measuring Skills for the 21st Century*

Research Council, *Psychological assessment in South Africa:* ***A needs analysis: The test use patterns and needs of psychological assessment practitioners***: Final Report, July 2004 Retrieved October 15, 2012 from http://en.wikipedia.org/wiki/Validity (statistics)

Search to Research Developing Critical Thinking Web Research Skills, ***Developing Critical Thinking Through Web Research skills*** **4937-4**
http://downhead.microsoft.com/downhead/A/6/4/A64.E848

Sevilla , C. (1992). ***Research methods*** (Rev. Ed.). Quezon City: Rex publishing.

Shuttleworth, M. (2009). ***Test-Retest Reliability.*** Retrieved October 15, 2013 from www.eplorable.com/test-retest-reliability

Stanley J., Hopkins K. ***Educational and Psychological Measurement and Evaluation***, New Jersey: Englewood Cliffs, 1981

Sternberg, R. J. (1986). ***Critical thinking: Its nature, measurement, and improvement.*** National Institute of Education. Retrieved October 14, 2012 from http://eric.ed.gov/PDFS/ED272882.pdf.

Sugrue, B. (1994). ***Specifications for the Design of Problem Solving Assessments in Science***(CSE Technical Report 387). Los Angeles: CRESST/University of California.

Sugrue, B. (1995). ***A theory-based framework for assessing domain-specific problem solving ability and Educational Measurement: Issues and Practices***. University of Cambridge.

Thayer-Bacon, B. J. (2000). ***Transforming critical thinking:*** *Thinking constructively*. New York, NY: Teachers College Press.

Turner, J. C. (1995). ***The influence of classroom contexts on young children's motivation for literacy.*** Reading Research Quarterly.

Tyler, Ralph W. ***Trends in Student Personel Work***.Minneapolis University of Minnesota, 1969

U.S.A. American Education Program, (1988).***The effect of gifted and talented programs on academic self- concept.*** Retrieved August 14,2012 from http://psycnet.apa.org/journals/amp/58/5/364/

Washington, DC: Education Sector. **Measuring Skills for the 21st Century**. Retrieved August 17, 2012 from http://www.educationsector.org/usr doc/MeasuringSkills .pdf.

Watson_Glaser (2002). ***Critical Thinking Appraisal***. Copyright by. Pearson Education,Inc. 2012

Wikipedia Foundation,Inc.(2013).**Test,Assessment**.Retrieved October 14,2012 from http://en,wikipedia.org/wiki./Test

Willingham, D. T. (2007). ***Critical thinking: Why is it so hard to teach?*** American Educator.

Woolfolk, Anita. ***Educational Psychology*** Eighth Edition, United States of America A Pearson Education Co., 2001.

A P P E N D I C E S

APPENDIX A

Table of Specifications
Critical Thinking Skill Test (CTST)

Critical Thinking Skills (Table 9.7)	No. of Times each Critical Thinking is used in the problem (Test)		
Content	Frequency	%	Item number
A. Defining and Clarifying the Problem			
1. identify central issues or problems	25	25	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25
2. Compare similarities and differences	15	15.15	2,3,4,5,6,7,16,17,18,19,20,22,23,24,25
3. Determine which information is relevant	23	23.23	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,20,22,23,24,25
4. Formulate Appropriate questions	9	.9	3,7,16,19,21,22,23,24,25
B. Judging Information Related to the Problem			
5. Check Consistency	1	1	18
C. Solving problems/ drawing Conclusions			
6. Recognize the adequacy of data	22	22.22	3,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25
7. Predict probable consequences	4	4	22,23,24,25
TOTAL	99	100%	

Appendix B
Letter Of Request (To the Principal)

Mrs. Leonila M. Buitizon
Santa Rosa elementary school Central-I
Santa Rosa City
Laguna

Dear Madam:

I am presently having my Seminar paper entitled "Development And Evaluation of A Critical Thinking Test In Mathematics For Grade III" for my Graduate Diploma in Master of Arts in Teaching Mathematics.

In this regard, May I request permission from your good office to administer a critical thinking test in mathematics for Grade-III pupils for my study.

Your kind and favorable action on this request is a gesture of your deep concern for the improvement of mathematics instruction.

Respectfully yours,

ESPERANZA M. BATITIS
Researcher

Approved:

MRS. LEONILA M. BUITIZON
Principal IV
Santa Rosa elementary School Central-I
Date Signed: September 12, 2012

Noted :

DR. ELISA S. BACCAY
Adviser

Appendix B
Letters to the Evaluator

To:

_____
Department of Psychology
Division of Santa Rosa City
Santa Rosa City, Laguna

Dear Madam:

I am presently having my special project entitled "DEVELOPMENT AND EVALUATION OF A CRITICAL THINKING TEST IN MATHEMATICS FOR GRADE III "

In this regard, may I request for your permission to evaluate my twenty-five item test instrument on selected topics for math III

Thank you very much.

Respectfully yours,

Esperanza M. Batitis
Researcher

Noted :

Dr.Elisa S. Baccay
Adviser

To:

_____
Department of Mathematics
Philippine Normal University
Taft Avenue,Manila

Dear Sir:

I am presently having my special project entitled "DEVELOPMENT AND EVALUATION OF A CRITICAL THINKING TEST IN MATHEMATICS FOR GRADE III "

In this regard, may I request for your permission to evaluate my twenty-five item test instrument on selected topics for math III

Thank you very much.

Respectfully yours,

Esperanza M. Batitis
Researcher

Noted :

Dr.Elisa S. Baccay
Adviser

APPENDIX C

Experts' Evaluation Instrument

The following items are used to evaluate the different critical thinking skill components of the twenty-five item test.

Please check the column that best describe your assessment of the test.

The meaning of the numbers are:

[1] Strongly disagree [3] Agree

[2] Disagree [4] Strongly Agree

Content	4	3	2	1
1. The item help young people develop capacities for independent judgement.				
2. The test items improve pupils abilities to think their way through content, problems and issues.				
3. The items used disciplined skill in reasoning.				
4. The items are progressive and create change in pupil overtime.				
5. The item states, that a pupil is in creating and exploring alternatives.				
6. The item measure the essential dimensions of critical thinking like the analysis of thought, the assessment of thought and the skills and abilities of thought.				
7. The test items can answer the question cues (aid in the stablishment and encouragement of critical thinking) of Bloom's taxonomy like describe,interpret, contrast, associate, and demonstrate.				
8. The item exhibits previously learned material by recalling facts, terms, basic concepts and answers.				

9. The items presents validity of ideas or quality of work based on a set of criteria. 10.The item depends opinion by making judgements about information. 11. The item propose alternative solutions. 12. The item makes inferences and finding evidences to support generalizations. 13. The items solve problems by applying required knowledge, facts, techniques amd rules in a different way. 14. The item gives descriptions and stating main ideas. 15. The items demonstrate undertaking of facts and ideas by organizing, comparing, translating and interpreting.				

Comments / Suggestions:

_____

_____

_____

APPENDIX D
Validation Instrument

The following items are used to evaluate the different critical thinking skill components of the twenty-five item test.

Please check the column that best describe your assessment of the test.

Content	Appropriate	Not Appropriate	Remarks
1. The item help young people develope capacities for independent judgement.			
2. The test items improve pupils abilities to think their way through content, problems and issues.			
3. The items used disciplined skill in reasoning.			
4. The items are progressive and create change in pupil overtime.			
5. The item states, that a pupil is in creating and exploring alternatives.			
6. The item measure the essential dimensions of critical thinking like the analysis of thought, the assessment of thought and the skills and abilities of thought.			
7. The test items can answer the question cues (aid in the stablishment and encouragement of critical thinking) of Bloom's			

taxonomy like describe, interpret, contrast, associate, and demonstrate 8. The item exhibits previously learned material by recalling facts, terms, basic concepts and answers. 9. The items presents validity of ideas or quality of work based on a set of criteria. 10.The item depends opinion by making judgements about information. 11. The item propose alternative solutions. 12. The item makes inferences and finding evidences to support generalizations. 13. The items solve problems by applying required knowledge, facts, techniques amd rules in a different way. 14. The item gives descriptions and stating main ideas. 15. The items demonstrate undertaking of facts and ideas by organizing, comparing, translating and interpreting.			

APPENDIX E

Developed Critical Thinking Test

Critical Thinking Test for Mathematics 3

_____1. Two dozen is ¼ of what number?

_____2. A box of milk usually contains 48 cans . How many cans of milk are there in 3 boxes?

_____3. If twice Nena's age increased by 8 is 26. How old is Nena?

_____4. A cooking class baked 95 cookies. They served 2/5 of the cookies to visitors.How many cookies were left?

_____5. How much more is the product of 8 x 9 than the sum of 8 and 9 ?

_____6. What is N in the sentence 15 + 65 + N = 15 x 12?

_____7. In a line, the only boy with glasses is 8th from the back and 11th from the front. How many pupils are in the line?

_____8. Gina can save P25 from her weekly allowance. In how many weeks can she save P500?

_____9. Because of lack of books, twelve books are shared by 20 pupils. What is the ratio of books to pupils in lowest terms?

_____10. There are 18 slices of bread in a loaf. How many 2-slices of sandwiches can be made from 4 loaves of bread?

_____11. The longer side of a rectangle is twice the shorter side. If the perimeter 36 dm, how long is each of the shorter sides?

_____12. Jim uses his bicycle for 90 km in 4 weeks to go to work. How far will he go in 7 weeks?

_____13. In a contest, Rica won P5,000. She gave ¼ of the money to her mother. If the mother spent half of what she received for groceries, how much did she spend?

_____14. Long-stemmed roses cost P180 a dozen. How much would it cost Nora to buy 3 ½ dozens?

_____15. What two numbers have a sum of 21 and a product of 108?

_____16. If you skip count by 7 and say 11 at the first skip, what will you say at the 11th skip?

_____17. Mark has three times as much money as his younger sister. If together they have P1460, How much money does Mark have?

_____18. Multiply the 12th multiple of 6 by the prime number between 61 and 71, and divide the product by 2, what is the result?

_____19. When I open my Mathematics book, there are two pages which face me. If the product of the two page numbers is 156, what are the two page numbers?

_____20. I am thinking of a number. If I add 12 to it and divide the result by 2, I get 10. What is the number?

_____21. How many numbers between 20 to 60 are divisible by 3?

_____22. What is N if 17, 843 = 17,000 + N + 40 + 3?

_____23. What is N if [N] + 5 [ ] x 3 [ ] ÷ 6 [ ] – 5 = [11]?

_____24. What is the smallest 4-digit number with no digits the same?

_____25. What is the biggest number you can form using 0, 5, 8, 7?

Appendix F

Curriculum Vitae

ESPERANZA M. BATITIS

L1 B1 Ambrocia Village Sta. Rosa City
Contact No. 09182432077/09478907678
E-mail Address: espiebatitis@yahoo.com.ph

SKILLS

- Teaching Skills
- Good Communication Skills
- Typing Skills
- Establishing rapport with parents and colleagues

PERSONAL DATA

Age	: 49 years old
Birthday	: August 26, 1963
Height	: 4'10"
Religion	: Roman Catholic
Civil Status	: Married

EDUCATIONAL ATTAINMENT

2009 – Present PHILIPPINE NORMAL UNIVERSITY
Master of Arts in Teaching Mathematics

1992 – 1995 LAGUNA NORTHWESTERN COLLEGE
Bachelor of Science in Elementary Education

1980 – 1984 PHILIPPINE WOMEN'S UNIVERSITY
Bachelor of Science in Business Administration

AFFILIATIONS

- Couples for Christ
  Facilitator: 1998-present
  Sta. Rosa Delima Parish Church

EXPERIENCE

- **Pre-School Teacher**
  Community Based Pre-School
  Santa Rosa Elementary School
  1990-1995

- **Grade I – Teacher**
  Santa Rosa Elementary School Central I
  1995-present

- Trainer of Grade I pupils of Central-I for MetroBank –MTAP-DepEd Math Challenge.

- Grade Chairman of Grade I Teachers at SRES Central-I.

- Treasurer of the SRES Central-I Teacher's Club.

- Working committee during the RACES Camp for Public and Private Elementary Schools.

- Facilitator and Lecturer in some teaching related trainings.

- Chairman of feeding program of Santa Rosa Elementary School