

Reliability of the Spanish Version of the Utrecht Early Mathematical Competence Test (Scale A)

Maria Elizabeth UNDERDAL



Masters' thesis
Master of Philosophy in Special Needs Education
Department of Special Needs Education
Faculty of Educational Sciences

UNIVERSITY OF OSLO

Spring 2012

Reliability of the Spanish Version of the Utrecht Early Mathematical Competence Test (Scale A).

© Maria Elizabeth Underdal

2012

Reliability of the Spanish Version of the Utrecht Early Mathematical Competence Test
(Scale A).

Maria Elizabeth Underdal

<http://www.duo.uio.no/>

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

This thesis is a quantitative research focused on the measurement of the Reliability of the Spanish Version of the Utrecht Early Mathematical Competence Test (Scale A).

The Utrecht Early Mathematical Competence Test was constructed in 1994 by Hans van Luit; Bernadette van de Rijt & Albèr H. Pennings, in Netherlands. Initially it had a pool of 120 items. Currently, it is composed by two scales (A and B) each consisting of 40 items.

The purpose of this test is to assess the developmental level of early mathematical competence in children ages 4 to 7 years, by evaluating eight mathematical concepts:

Comparison of quantitative and qualitative characteristics of objects; classification of objects in class or subclass; correspondence one to one relation; seriation of objects in class or subclass based on criteria; counting words forward and backward; structured counting; resultative counting and general knowledge of numbers.

As a primary teacher graduated from the Pontifical University of Ecuador, I have always been concerned about mathematical learning disabilities and I believe that the use of an appropriate assessment tool that helps us teachers to know the mathematical development of kindergarten students can be crucial in preventing future learning problems or inferior performance later. It can also contribute in a significant way to special educational planning and decision-making, for example, in terms of more accurate evaluation and diagnosing, in establishing supportive social networks, and in the designing of appropriate learning materials.

Among some instruments used in Ecuador to assess the mathematical skills in children we can find the Spanish version of the Utrecht Early Mathematical Competence Test (Scale A). This version was developed in the University of Cadiz, Spain. Considering that there have been no studies about the reliability of the results obtained in the application of this test in Ecuador, the aim of my project will be to measure the reliability of this version when it is used in Ecuadorian kindergartens.

The research is focused on reliability because it reflects the ability of an instrument to perform and maintain its functions in routine. A low punctuation of the reliability of a test can be a sign that the items are ambiguous or that the standard conditions during the

administration of the test are failing. In other words, measuring the reliability of the test will show us the consistency of this measure in the Ecuadorian environment.

Between the several general classes of reliability estimates, two types of reliability were estimated for this research: The test-retest and the internal consistency. The study was carried out on a sample of one hundred Ecuadorian children, between four and five years old. The children were students of 3 different kindergartens in Quito, which are well known for their good academic level and they are ranked on a medium-high economical level.

After applying the statistical techniques to the data collected through the SPSS program, the statistical estimations showed a high degree for both, test-retest reliability and internal consistency reliability. These results of the empirical studies confirmed that the Spanish version of the Utrecht Early mathematical competence is a reliable assessment method for examining the numerical abilities of Ecuadorian preschool children.

Foreword

This thesis was written for my Master degree in Special Needs Education at the University of Oslo during the time period from Autumn 2011 until Spring 2012, under the teaching supervision of Professor Peer Møller Sørensen. The collection of data of this research was done in Quito, Ecuador between 01.09.2011-01.01.2012. The rest of the research and writing of the dissertation took place in Oslo, Norway.

The purpose of this research is to measure the reliability of the Spanish Version of the Utrecht Early Mathematical Competence Test (Scale A) within the Ecuadorian Kindergartens.

I would like to thank my supervisor Professor Peer Møller Sørensen for his suggestions, encouragements and guidance in writing the thesis and approaching the different challenges during the thesis.

The field work done in Ecuador was extensive, and it is not possible to thank all the persons involved by name, I hope that all of the administrators, teachers, children and parents accept my thanks in this form.

Finally I would like to thank my parents for their constant support during the time I studied my master degree.

Table of Contents

Chapter 1: Introduction to the study	1
1.1 Framework of the research.....	1
1.1.1 Statement of the Research problem and justification.....	2
1.2 Research design and method.....	3
1.3 Prior researches and Literature review.....	5
1.4 Ethical implications.....	11
Chapter 2: Learning disabilities and mathematical learning problems.....	14
2.1 Learning disabilities definition and different approaches to this concept.....	14
2.2 Mathematical Learning Disability Definition.....	18
2.3 Different types of deficits in mathematical learning disabilities and causes.....	20
Chapter 3: The development of Psychometrics.....	27
3.1 A historical review of Psychometrics.....	27
3.2 Psychometrics' main theory.....	28
3.3 Standards for Educational and Psychological Testing	30
Chapter 4: The Utrecht Early Mathematical Competence Test.....	36
4.1 Structure of the test.....	37
4.1.1 Concepts of the test.....	37
4.1.2 Items and instructions.....	39
4.1.3 Norms and Interpretations of the results.....	44
4.2 Background of the test.....	47

Chapter 5: Statistical estimations, analysis and conclusions.....	51
5.1 Reliability.....	51
5.2 Statistical Estimations.....	54
5.3 Discussion.....	57
5.4 Conclusion.....	60
Literatur Reference List.....	64
Appendix.....	68

Statistical tables and figures.

Table 1: Age groups.....	45
Table 2: Norm Table for Version A.....	46
Table 3: Cronbachs alpha values for interpreting internal consistency.....	53
Table 4: Internal consistency for all items. Administration <i>a</i> and <i>b</i>	54
Table 5: Internal consistency for each subscale. Administration <i>a</i>	55
Table 6: Internal consistency for each subscale. Administration <i>b</i>	55
Table 7: Test-retest for each subscale. Administration <i>a</i> and <i>b</i>	57
Figure1: Pearson Correlation.....	56
Figure 2: Frequency distribution of the raw scores in administration <i>a</i>	Appendix.
Figure 3: Frequency distribution of the raw scores in administration <i>b</i>	Appendix

Chapter 1: Introduction to the study.

In this chapter an overview of the theoretical framework, the methods and design of the research is described. It also presents the statement of the research problem. The chapter concludes with an illustration of prior studies related to this research, the literature review and the ethical implications involved in this work.

1. 1 Framework of the Research.

The focus of this research is the use of the Utrecht early mathematical competence test as a reliable instrument to assess the early mathematical skills of children. These early mathematical concepts that children develop through their first years of life, represent the base for their future learning of mathematics at school.

A key assumption that underlies cognitivism is that prior knowledge plays an important role in learning; therefore part of the framework of this research is based on a cognitivist learning approach.

Cognitive Learning Theory also implies that the different processes concerning learning can be explained by analyzing the mental processes first. It posits that with effective cognitive processes, learning is easier and new information can be stored in the memory for a long time. On the other hand, ineffective cognitive processes result into learning difficulties that can be seen anytime during the lifetime of an individual. This idea is supported by the eclectic psychologist Reuven Feuerstein. His approach relies on a constructive constructivist paradigm.

Feuerstein claims that learning problems may be prevented through early, developmentally appropriate, intervention, and since one of the goals of the Utrecht early mathematical test is to prevent future learning mathematical problems, the theory of Structural Cognitive Modifiability (SCM) developed by Feuerstein constitutes an important component of this framework.

The SCM views the human organism as open, adaptive and amenable for change. The aim of this approach is to modify the individual, emphasizing autonomous and self-regulated change. Intelligence is viewed as a propensity of the organism to modify itself when confronted with the need to do so. It involves the capacity of the individual to be modified by learning and the

ability to use whatever modification has occurred for future adjustments. Intelligence is defined as a changeable state rather than an immutable trait. Cognition thus plays a central role in human modifiability. Many behavioral and emotional conditions may become modified through cognitive intervention.

Mediated Learning Experience is a proximal factor of human modifiability, which can moderate the influence of such distal factors as genetic predisposition, organic impairment, or educational deprivation.

As a final point, it is important to mention that the concept of early intervention plays also an important role in this framework. Child development research has established that the rate of human learning and development is most rapid in the preschool years. Timing of early intervention becomes particularly important when a child runs the risk of missing an opportunity to learn during a state of maximum readiness. If the most teachable moments or stages of greatest readiness are not taken advantage of, a child may have difficulty learning a particular skill at a later time. Thus, through the correct use of the Utrecht Early mathematical competence test, mathematical challenges faced in the early childhood can be identified and the teachers can work on these weaknesses or deficits in order to prevent future math disorders at school.

1.1.1 Statement of the Research problem and justification.

Learning disabilities in the area of mathematics are usually complex and require intervention by skillful teachers. This intervention should start at early childhood and in order to make this possible it is required to have an assessment instrument to evaluate the mathematical development of kindergarten students.

An assessment instrument that determine the early mathematical skills achieved by a child can contribute in a significant way to special educational planning like for example having more accurate evaluation and diagnosing processes, establishing supportive social networks, and designing appropriate learning materials.

Among some instruments used in Ecuador to assess the mathematical skills in children we can find the Spanish version of the EMCT (Scale A), developed in the University of Navarra.

Since there have been no studies about the reliability of the results obtained in the application of this test in Ecuador, the aim of my project will be to measure the reliability of this version when it is used in Ecuadorian kindergartens.

The project was focused on reliability because it reflects the ability of an instrument to perform and maintain its functions in routine. A low punctuation of the reliability of a test can be a sign that the items are ambiguous or that the standard conditions during the administration of the test are failing. In other words, measuring the reliability of the test will show us the consistency of this measure in the Ecuadorian environment

1.2 Research design and method.

The research in the present dissertation is a quantitative research, conducted through a non-experimental descriptive design. This means that the phenomenon was studied as it exists, without any intervention or treatment.

The data was gathered by administering the EMTC test to a sample of one hundred Ecuadorian kindergartens children. The sample was drawn by convenience from an accessible population. The accessible population constituted the neighborhoods in northern Quito (Ecuador's capital) and the children were selected from three Kindergartens from that area of the city. The age range of the sample is from 4 to 5 years. The age range of the sample in this research could have been from 4 to 7 years since the Utrecht Early Mathematical Competence Test cover this age range, however due to my personal interest in the development of mathematical concepts and skills in the first years of childhood, I decided to focus in the performance of children from 4 to 5 years old.

As it is mention before, the sample was selected according convenient and specific criteria determined by the purpose of the research: the location of the sample was selected because of the easy accessibility that I as the researcher had to this area of the city. The size of the sample, one hundred children, was determined by the available time that I had to administer the test. The test must be administered individually, and the administration of it takes around fifteen to twenty minutes per child. The test was administered twice to the same children following the requirements of the test re-test method that will be described later in the forth chapter. It took two weeks to administer the test to the whole sample for the first time in

October 2011, and then it took approximately also two weeks to administer it to the entire sample for the second time in December 2011. Lastly, it should be mentioned that the 3 institutes where the children study were selected for the following reasons:

- The three kindergartens belong to the same socio-economic level (upper middle class).
- The three Institutes are well known for having a high standard of education. This fact reduced the probability of teaching challenges interfering with the scores obtained by the children.
- Personal familiarity with the administrators who need to approve data collection.

Concerning the research method used on this study, it was a statistical analysis method the one that was employed to obtain the information from the data. The processing of statistical data was done by the SPSS Software. As the title of this dissertation describes, the objective of this study was to measure the reliability of the test.

There are different classes of reliability estimates. In this project the test-retest reliability and the internal consistency reliability. These estimates were chosen because they are best used for aspects that are stable over time, such as intelligence.

After the statistical analysis, the results were interpreted according to the level of stability that they showed, and correspondent conclusions were drawn.

Regarding the validity and reliability of the study, many aspects were taken into account in order to avoid threats against these important features of the research. To ensure internal validity in this study, all situational specifics like time, location, lighting, noise, timing etc, were carefully selected and reproduced in the same way during each administration of the test. This is very relevant in order to keep away from extraneous variables that may bear any effect on the behavior of the subject being studied. Also, considering the threat of differential attrition, the sample was selected with the children that the core teachers believe to be the less likely to miss classes, according to their previous experience with these groups of students.

On the subject of external validity, it is important to remember that since it is the degree to which research findings can be applied to the real world beyond the controlled setting of the research, the focus of this type of validity is the generalization. In view of the fact that the sample for this study was selected by convenience, then the generalization of the obtained

inferences will be done to a sample that meet the same features of the sample used for this research. Therefore a very precise description of the main features of the sample in this study will was presented before in this chapter.

Now, focusing on the reliability of this study, we would like to emphasize that it is an essential pre-requisite for validity. It is possible to have a reliable measure that is not valid; however a valid measure must also be reliable. In order to enhance the reliability of the present study, the raw scores of each test and the raw data of all research findings will be kept, plus a clear precise description of the research methodology from sample selection and other research aids have been well described in this section so that other researchers will be able to reproduce the methodology exactly.

At last, it is pertinent to mention that all the raw data (tests sheets), hand computations and computer printouts used in this research will be kept in order to permit the use of this information in future investigations or in case of any revision or reconstruction of the present study.

1.3 Prior Researches and Literature Review.

Over the last several decades, important advances have been made in understanding the deficits that underlie many learning disabilities. Important researches have been conducted in the area of reading disabilities and that phenomenon now seems to be better understood. An understanding of disabilities in mathematics, seem somewhat more elusive. One difficulty is the complexity of mathematics. Each domain of mathematics (algebra, geometry, trigonometry) is quite intricate and contains multiple sub-domains. In addition, children with disabilities in mathematics have unique patterns of strengths and weaknesses; there is no one mathematics disability (Berch & Mazzozo, 2007).

However, in the recent years, we have seen a major shift within the fields of mathematics education from a mainly psychological and pedagogical perspective towards one that encompasses the historical, cultural social and political contexts of both mathematics and mathematics education. This multitude of factors is having an unprecedented impact on mathematics education and its research endeavors. This fact is reaffirmed by the emergence of international assessments like the Trends in International Mathematics and Science

Study (TIMSS). The TIMSS constitutes an assessment of the mathematics and science knowledge of 9–10 and 13–14 year old (Year 5 and Year 9 or fourth grade and eighth grade) students around the world. TIMSS was developed by the International Association for the Evaluation of Educational Achievement (IEA) to allow participating nations to compare students' educational achievement across borders. TIMSS was first administered in 1995, and every 4 years thereafter. In 2007, 48 countries participated.

So, there is a worldwide increasing interest about the mathematical development in children and regarding specifically to the mathematical skills in early childhood, different theories and concepts have been raised since the pioneer work of Piaget in this matter. In a chronological way, some of the main ideas developed by psychologists and other researchers concerning mathematical learning in children are presented below:

From the pioneer studies of Piaget and Szeminska (1941), it was considered that the development of logical thinking is the basis for the development of number and arithmetic skills in children (Dehaene, 2001). According to this approach, mathematical development is linked to the development of logical thinking, for example, we said that a child understand the concept of number at the time that he or she controls the principles of logic and the use of inferences. Also the maintenance operation plays an important role in the overall Piagetian theory. The numbers would not be understandable if they were not identical to themselves no matter any of their apparent transformations. In short, the Piagetian model has had an enormous influence on the teaching and learning of mathematics. Also the model has been used as a theoretical framework for understanding the dyscalculia.

The criticism of the Piagetian model have been varied and although some of them are unfounded, other experimental work come to question the number operative model advocated by Piaget, considering that the model provides an incomplete explanation of numerical skills in children (Barouillet & Camos, 2006).

An alternative approach argues that there is no clear relationship between the development of number and logic operations. On the contrary, this approach argues that the understanding of number develops gradually through the experiences of the child counting (Gelman & Gallistel, 1978; Barouillet and Camos, 2006; Bideaud & Lehalle, 2002).

According to this framework, the count is seen as a more complex notion, and not just a rote recitation of the oral-numerical string that goes from concrete levels to more abstract

levels. The initiation of children in the world of number is given in the context of aging, so that the interactions that occur within the family are related to the development of the numerical understanding: songs with numbers, rhymes, games, birthdays, etc. Therefore, in early development, children become familiar with numbers in many different ways.

This approach has allowed us to identify accurately the progression and development of mathematical knowledge between two and seven years of age (Clarke and Cheeseman, 2007). The conclusions of these studies assume that in addition to those Piagetian logical operations, various counting skills are also important for the development of number and thus learning the conventional numbering system would begin in early childhood with the acquisition of the verbal sequence the numeric string.

It is largely based on the studies mentioned above, and from an interactionist perspective, that Van de Rijt, Van Luit Rijt, assumes the idea that the Piagetian and counting operations processes do not need to be separated and that together contribute to the development of the number; and based on this idea, they constructed, in the 90s decade, the Utrecht Early Mathematical Competence Test.

The EMTC has been used in many European researches to validate longitudinal researches as well as to use it with early age students, special learning needs and with children that have difficulties on mathematics. Some examples of these investigations are:

- A research of the EMCT was carried out on 127 third year Spanish kindergarten school children by the University of Cadiz in Spain, 2008. The students from the sample were from three different schools of a 135.000 population town and they were ranked on a medium or medium-high economical level. The objective of this research was to know the mathematical development of kindergarten students concerning how they make relate concepts and count through an evaluation.
- During the period 2008-2009 a research using the UEMTC as the main assessment instrument of children mathematical skills, was conducted by the University of Helsinki, the Utrecht University, and the Chinese Academy of Science. This study examined the influence of nationality, age and gender on Chinese (N:130) and Finnish (N:203) pre-schoolers' number sense. Two highly correlated aspects of number sense were extracted: one reflecting the children's ability to organize and compare quantities

(relational skills), and another pertaining to their ability to operate with number-word sequence (counting skills). The results showed a significant age-related gain in both aspects of number sense, whereas no gender differences were found. With respect to counting skills, the Chinese children outperformed the Finnish children irrespective of age, whereas in relation to relational skills, this was true only among the older children. Differences in language, teaching and cultural ethos are considered as alternative explanations for the findings.

- In 2009-2010 a longitudinal study from the University of Jyväskylä and the University of Helsinki in Finland, examined through the administration of the EMCT how children's early numeracy assessed in kindergarten predicts their mathematical performance in the first grade, after controlling for the effects of age, gender, and parents' education. The participants were 212 Finnish children (107 girls and 105 boys). At the time of the first assessment the mean age was six years, and the second assessment was conducted one year later. The results demonstrate that the acquisition of counting and relational skills before formal schooling are predictive of the acquisition of basic arithmetical skills and overall mathematical performance in grade one, above and beyond the effects of demographic factors.

As we can see, the question of how children acquire the necessary skills to learn mathematics is becoming a very interesting subject of research among psychologists, pedagogues and other professionals all over the world.

In the next section, a brief literature review, holding the main books and articles that were a valuable source of information for the framework of this study with respect to the concepts of learning processes in children, mathematical disorders and reliability, will be presented.

Literature Review on *learning processes* and *mathematical disorders*:

1. Dockrell, J. (1998), Development of mathematical skills. England.

This book describes studies of number awareness in early childhood and presents an overview of the basic principles in the acquisition of mathematical knowledge. It explores the reasons of individual differences in mathematical skills and the influence of the cultural context in the mathematical learning process.

2. Berch, D. &Mazzozo, M. (2007). Why is it Math so hard for some Children?: The nature and origins of Mathematical Learning Difficulties and Disabilities.

This book analyses the origins of mathematical learning difficulties. It describes what is behind the students' difficulty from diverse perspectives from fields such as special education, educational psychology, cognitive neuroscience and behavioral genetics.

3. Gellistal, R. C., & Gelman, R.; (1978). The child's understanding of number. USA, President and Fellows of Harvard College.

Gellistal, R. C., & Gelman, R examine the logic of preschool children. They describe cognitive skills of preschool infants focusing in the child's understanding of number. They present different paradigms and hypothesis that have been raised about this aspect from different psychological perspectives.

4. Graham, A., Lyn, D., & Bussi, M., Handbook of International Research in Mathematics, (2008). USA, Taylor and Francis.

This handbook presents an overview of international research in mathematics education. It also collects some ideas for making more productive the mathematics education research, especially regarding the policies and practices.

5. Romberg, T., (1992). *Mathematics Assessment and Evaluation*. USA. States University of New York.

This book shares a set of papers written on issues surrounding mathematics test and their influence on school mathematics. It also describes trends in evaluation and assessment that show the disparity between what is possible and what is in fact achieved in the area of mathematical assessment.

6. Schweingruber, H., Woods, T., & Cross, C., (2009). *Mathematics learning in early childhood*. USA, National Academic Press.

This book examines the cognitive foundations for early mathematics learning and it describes some developmental variations, sociocultural influences and difficulties in the mathematical learning process.

7. Wong, B., Graham L., Hoskyn, M., & Berman, J. (2008). *The ABCs of Learning Disabilities*. USA, Elsevier Academic Press, Second Edition.

Wong, B., Graham L., Hoskyn, M., & Berman, J. present an overview of the understanding of learning disabilities through history, both in Europe and America. They also describe some different definitions of learning disabilities and the achievements in the learning disabilities field.

Literature Review on *Reliability*

1. Sirkin, M. (2006). *Statistics for Social Science*. USA, Sage Publications.

This book describes different statistical and data analysis techniques for social science researches. It provides clear examples where these techniques should be used, and it explains the advantages and disadvantages of diverse statistical methods.

2. Traub, R. (1994). *Reliability for the Social Sciences: Theory and applications*. USA, Sage Publications.

This book presents an introduction to the theory and application of classical approaches to measurement reliability. It also presents a discussion of the concept and meaning of reliability in everyday life and in social science.

3. Zeller, R., Carmines, E. (1980). *Measurement in the social sciences: the link between theory and data*. USA, Cambridge University Press.

Chapter 3 of this book describes the different types of reliability. It provides a detailed explanation of the use of Cronbachs Alpha in social researches including the possible problems and limitations that this measurement may imply.

1.4 Ethical implications.

Having a mathematical test as the central point of a research implies some ethical issues that must be taken into account.

One such issue is the right to privacy. The concepts of individual rights and privacy are an essential part of our society. The Ethical Principles assert individual rights to privacy and confidentiality.

A common understanding about the right of privacy is:

“The right to privacy is our right to keep a domain around us, which includes all those things that are part of us, such as our body, home, thoughts, feelings, secrets and identity. The right to privacy gives us the ability to choose which parts in this domain can be accessed by others, and to control the extent, manner and timing of the use of those parts we choose to disclose”.
(Yael Onn, 2005)

So in order to apply the Utrecht Early Mathematical Competence Test (EMTC) to the sample in this research, it was important to obtain the correspondent parental permission. For getting this permission it was necessary to give to the parents or guardians of the children a complete and clear explanation regarding how exactly the results of the test were going to be used and what do the results mean. This explanation was provided by the each one of the directors from

the three kindergartens. The directors used a letter to inform the parents about the research and the parents sign the letter to confirm their permission. These complete explanations are commonly known as informed consent and they were conveyed in such a way that was straight-forward and easy for parents to understand.

Another issue regarding the privacy aspect of the individuals involved in the test is the handling and storage of the test results and personal data. Once test results are held on computer file, then the subject of them is legally entitled to access. It is also important to give consideration as to who else may have access to them. Two criteria usually used are that they should be available to those only with a genuine interest and/or those appropriately trained in interpretation.

There is also the question of where and how such test data should be held. Test results, like all personal information, should be stored with due regard to confidentiality. Access should be restricted to those with a need to know and in accordance with what has been agreed with the parents and kindergarten. Persons who are untrained should not be allowed access to raw data from tests but only to clearly described interpretations.

It is also important to remember that individuals do change and develop and so psychometric data can become less accurate over time. Test scores should therefore not be kept on file indefinitely. The time period for which scores are valid will differ depending on the nature of the measures and the particular use made of them. While the best tests can provide well-constructed measures of aspects of individuals at a point in time, this does not preclude a degree of subsequent change and development, especially with younger candidates and in the personality domain. As a rule of thumb, test results more than 18 months old should not be used. (Yael Onn, 2005)

Finally, it is necessary to consider that the results of a test that measures intellectual and cognitive skills can be interpreted as diagnosis labels. Traditional diagnostic labels have served several purposes: They have helped professionals keep track of the different kind of problems children are having and helped researchers study the causes for those problems. But diagnostic labels also have important limitations. When we try to group different individuals under a large category of what they appear to have in common, we risk grouping together children who are actually quite different from one another. One example of this is the attention deficit hyperactivity disorder (ADHD) label. The ADHD diagnosis focuses on the

similarity among children who are inattentive, and maybe unable to concentrate well enough to follow directions. By settling for the label ADHD to explain the behavior of such children, we underemphasize many important differences among them. One child may be inattentive because he can't plan his actions well. Another may have great trouble processing incoming information. Yet another may be oversensitive to sound and, when confused, becomes inattentive and disruptive.

Therefore there might be parents who may not approve to use the results of their children's test in the research since they consider that low results on the Utrecht Early Mathematical Competence Test is a diagnosis which predicts a poor performance of their children in mathematics during their whole life or they might believe that low results in the test must necessarily mean that their children have some kind of mental disability.

Chapter 2: Learning disabilities and mathematical learning problems.

This chapter contains a brief description of the different perspectives from which the terms learning disabilities and mathematical learning disabilities have been defined. It presents some of the problems derived by the vagueness and ambiguity that surrounds the understanding of these terms.

The chapter concludes with a general depiction of the different types of mathematical learning problems and the possible causes for these disorders.

2.1 Learning disabilities definition and different approaches to this concept.

According to the definition proposed by the American institution, the National Joint Committee on Learning Disabilities (NJCLD) in 1990, the term learning disabilities refers to:

“A heterogeneous group of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, reading, writing, reasoning, or mathematical abilities. These disorders are intrinsic to the individual, presumed to be due to central nervous system dysfunction, and may occur across the life span. Problems in self-regulatory behaviors, social perception, and social interaction may exist with learning disabilities but do not by themselves constitute a learning disability. Although learning disabilities may occur concomitantly with other handicapping conditions (for example, sensory impairment, mental retardation, serious emotional disturbance), or with extrinsic influences (such as cultural differences, insufficient or inappropriate instruction), they are not the result of those conditions or influences”.

It is necessary to underline that as it usually happens with many social concepts, there is a wide range of issues and problems surrounding the terminology used in learning disability. This term can mean many things to different people.

Learning disability has been source of speculations, fear, and scientific enquiry for hundreds of years. It has been regarded in turn as an administrative, medical, eugenic, educational and social problem (Clarke & Cheeseman, 2007).

Therefore, in order to get a deeper understanding of what does the term learning disability involves, it is important to have an overview of the theoretical perspectives that are commonly associated with this term.

From a sociological perspective, the inability of a person with a learning disability to undertake functional activities is emphasized. The functional activities are understood as everyday experiences for most people. These are activities such as going to work, taking care of one-self, etc.

This sociological perception of learning disability has led to the emergence of two main images of individuals with learning disabilities:

- Persons with learning disabilities as sick persons.
- Persons with learning disabilities are developing persons.

The first image is frequently held by doctors and medical personnel, while the second is held by parents, psychologists and teachers.

As we can see, from a sociological point of view, it is the social role of the person with learning disabilities that constitutes the main idea of the concept, and the expectations of this role become to some degree limited by the images that the society constructs about people with learning disorders.

From a psychological perspective, we find different theoretical explanations to understand the concept of learning disability. For example, from a behavioral approach, the major focus of interest regarding learning disorders would be the individual problems and the identification of objectives for improving individuals' ability to perform observable behaviors that they are currently unable to do. In contrast, from other perspectives such as the interventionist approach, the observable behaviors are not the only important issue, but environmental aspects such as thoughts and feelings are important as well.

Between the ranges of concepts within the discipline of psychology, the "restricted developmental" perspective is considered as one of the best approaches to understand the nature of learning disabilities. The argument to defend this idea is that this approach is an integral component of empirically based behavior theory of human development.

From a medical perspective and regarding to western medicine practice which uses a deductive approach that follows a diagnostic path on search of an explanation for a disease, the main focus in understanding learning disabilities is the identification of the different disorders and the determination of their causes in order to prevent or ameliorate them. The medicalisation of the term learning disability, leads to the fact that learning disability and mental

illness are frequently viewed as the same. Therefore the medical advances, especially in pharmacology have been extrapolated to people with learning disabilities.

As we can see this medical approach has contributed to the understanding of some of the more biological dimensions of learning disability.

Finally, we can consider the definition of learning disability from a cultural anthropological perspective. Many anthropologists identified that learning disability is a culturally defined phenomenon and that therefore cultural differences make terms like “learning disability” very deceptive.

Anthropologists argue that there are numerous life-history accounts of people with learning disabilities which are clearly contextualized within the cultural fabric of the communities and the societies to which these people belong. These kinds of accounts, give professional careers and general public, valuable insights into the nature of learning disability and some of its temporal dimensions.

As we can observe the concept learning disability can be understood from different perspectives and the interpretation of its definition has resulted in a series of problems that have affected in a theoretical and practical way many issues related to this field.

It is important therefore to underline that in this thesis, the term learning disability is understood according to the definition provided by the National Joint Committee on Learning Disabilities, which has become worldwide used and that takes into account important considerations that other definitions don't, such as:

Heterogeneity: learning disabilities should be recognized as a general term referring to heterogeneous group of disorders which are realized as significant difficulties in the acquisition and use of one or more of the following functions: listening, speaking, reading, and mathematical abilities. Individuals with such disabilities also may evidence problems in their ability to self-regulate behaviors and demonstrate altered patterns of social perception and social interaction. The idea that these problems can exist with learning disabilities has been acknowledged by the NJCLD and is consistent with current research findings. The inclusion of this idea within the definition is, therefore, opportune and contributes to a better understanding of individuals with learning disabilities.

Extension of the applicability of the term: the use of “children” in the federal definition limits the applicability of the term ‘learning disabilities’ to individuals 0-21 years of age. This results in a failure to recognize the developmental nature of learning disabilities. Indeed, learning disabilities must be viewed as a challenge not only of the school years, but of early childhood and continuing into adult life.

Etiology of learning disabilities: the NJCLD urges that the disorders represented by the collective term “learning disabilities” are understood as intrinsic to the individual and that the basis of the disorders is presumed to be due to central nervous system dysfunction. Although the NJCLD supports the idea that failure to learn or to attain curricular expectations occurs for diverse reasons, learning disabilities have their basis in inherently altered processes of acquiring and using information. It is essential to understand this notion if one is to appreciate the resultant interaction between the learner and the learning environments. An understanding of this interaction facilitates the development of effective service delivery models and adaptive curriculum. This also leads to a clearer understanding of the ways in which individuals with learning disabilities may interact in a life-long social and cultural milieu.

Different cultural and linguistic backgrounds: it is essential to understand and recognize the learning disabilities as they might occur within the varying disability categories as well as different cultural and linguistic groups. Individuals within these groups frequently have received inappropriate educational assessment, planning, and instruction because they could not be identified as learning disabled. The NJCLD supports the idea that learning disabilities are not the primary and direct result of other disabilities and should not be so confused. However, the NJCLD notes specifically that learning disabilities may occur concomitantly with other disabilities.

The reason why the definition of learning disability developed by the NJCLD is one of the most broadly accepted is the detailed specification of the criteria that should identify this concept. This definition has been formally adopted as the official definition of learning disabilities by the following NJCLD member organizations:

American Speech-Language-Hearing Association, Council for Learning Disabilities, Division for Children with Communication Disorders, International Reading Association, National Association of School Psychologists, and Orton Dyslexia Society.

2.2 Mathematical learning disability definition.

As is the case with the definition of learning disabilities, there is much ambiguity around the concept of mathematical learning disability. In addition, there is not a consensus about the term to be used. Several authors refer to this disability using different terms: “disablement in mathematical problem solving”, “mathematics learning difficulty”, “mathematics learning retardation”, “mathematical learning deficiency”, “dyscalculia”, etc. (Desoete, 2007).

Most of the definitions of these terms are based on assumptions of average – or above average- ability IQ, normal sensory function, adequate educational opportunity, and absence of other developmental disorders and emotional disturbance.

Due to the vagueness of the existent definitions, no standards have been established by which to judge the absence or presence in learning disabilities in math. This dilemma also gets affected by the fact that these varieties of terms imply broad concepts and are used for a wide range of impairments, ranging from computations, to problem solving and words’ problems. The broadness of these terms has led to the search of a definition that implies more specific characteristics of people with mathematical learning disability which in turn has caused the use of the term dyscalculia a synonymous of any mathematical learning problem.

Another major problem in defining mathematical learning disabilities is the need of focus on identification of a set of key academic skills deficits that represents markers for one or more learning disabilities in math. This identification should proceed from a model identifying critical components of math proficiency. Unfortunately, the understanding of the numerical competencies that characterize math is not as well developed as the reading or writing competencies; and even less clear is whether there are academic skill deficits involving math reasoning that cannot be explained by difficulties with reading and language.

In developmental models of math, conceptual and procedural aspects of mathematical knowledge are assumed to be required for the performance of many mathematical tasks, and the development of mathematical skills emerges from the reciprocal nature of the relationship between conceptual and procedural knowledge. This raises fundamental questions about attempts to separate knowledge of mathematical concepts from mathematical computations in definitions of mathematical disabilities.

Given the difficulty in defining a group of academic skill deficits that identify individuals with mathematical learning disabilities, research has not advanced to a level that allows the identification of a set of core cognitive processes that underlie learning problems in math.

At the very least much will depend on the type of theoretical orientation and mathematical competencies that are used to identify the math learning disability.

Some standardized definitions and definitions given by renowned researches, describe the term of mathematical disability or mathematic disorder as:

“A structural disorder of mathematical abilities which has its origin in a genetic or congenital disorder of those parts of the brain that are the direct anatomic-physiological substrate of the maturation of the mathematical abilities adequate to age without a simultaneous disorder of general mental functions”. (Kosc, 1974)

“The low achievement of a person on a certain occasion which manifests itself as performance below standard of the age group of this person or below his own abilities as a consequence of inadequate , cognitive, affective, volitional, motor or sensory, etc. development. The cause of inadequate development may be of various kinds”. (Kline, 1980)

“A disorder in the ability to do or to learn mathematics, that is, difficulty in number conceptualization, understanding number relationships and difficulty in learning algorithms and applying them. It is an irregular impairment of ability”. (Sharma, 1986)

“A condition that affects the ability to acquire mathematical skills. Learners may have difficulties understanding simple number concepts, lack an intuitive grasp of numbers, and have problems learning number facts and procedures. Even if they produce a correct answer or use a correct method, they may do so mechanically and without confidence”. (McCloskey,1992)

“Difficulties in production or comprehension of quantities, numerical symbols, or basic arithmetic operations some are not consistent with the person's chronological age, educational opportunities, or intellectual abilities. The disturbance significantly interferes with academic achievement or activities of daily living that require these numerical skills”. (APA, DSM-IV)

The last definition described above is the one used by the American Psychiatric Association in the DSM-IV. This definition is under the term of mathematics disorders, however it has

now been proposed that the disorder should be called dyscalculia. If this occurs, the name change will happen in 2013 when the DSM-V (5) will be released, and will essentially change the way professionals view dyscalculia.

As we can see there is a variety of terms regarding mathematical learning disabilities, and there is no definite agreement on their use universally in the literature and many authors have used these terms interchangeably.

Throughout this work, in order to avoid ambiguity and trying to be consistent with the international use of this notion, the terms mathematical disabilities or mathematical disorders will be understood according to the definition and criteria given by the American Psychiatric Association.

2.3 Different types of deficits in mathematical learning disabilities and causes.

David Geary (1999) distinguishes five basic components involved in cognitive deficits in children with mathematical learning difficulties:

- Deficits in counting or other types of procedures:

A slow memorization of the counting sequence and a low speed in processing information, are the main problems in counting and related procedures. Since the basic strategy to solve the first addition and subtraction problems is based on counting procedures, students with these kind of difficulties will face challenges in mathematics from the early years of schooling.

- Deficits in the memorization of numerical facts:

One clear example of these kinds of difficulties is the memorization of multiplication tables. Children with this problem struggle with the long-term memory. They calculate using the basic counting and in most cases using their fingers to keep counting, but this may not be very useful when they need to calculate harder multiplications such as 9×8 .

For some students, this represents their only notable math learning difficulty and, in such cases, it is crucial not to hold them back "until they know their facts." Rather, they should be

allowed to use a pocket-size facts chart in order to proceed to more complex computation, applications, and problem-solving. As the students demonstrate speed and reliability in knowing a number fact, it can be removed from a personal chart. Addition and multiplication charts also can be used for subtraction and division respectively. Also, by blackening over each fact that has been mastered, overreliance on the chart is discouraged and motivation to learn another one is increased.

- Deficits in conceptual knowledge:

Some children are not able to make an adequate conceptualization of operations. Even though many of them actually bring to school a strong foundation of informal math understanding, they encounter trouble in connecting this knowledge base to the more formal procedures, language, and symbolic notation system of school math. The collision of their informal skills with school math is like a tuneful, rhythmic child experiencing written music as something different from what he/she already can do. In fact, it is quite a complex feat to map the new world of written math symbols onto the known world of quantities, actions and, at the same time to learn the peculiar language we use to talk about arithmetic. Students need many repeated experiences and many varieties of concrete materials to make these connections strong and stable.

Children with needs in conceptualization are usually able to solve problems if they work with any concrete referent such as fingers, materials or graphical representation, but without these resources, it is difficult to take the step of concrete situations in mathematical symbolization and to establish connections between some situations and others. However a good number of them do not show difficulties in areas such as geometry, and probability concepts or measure; their problems are often mainly with arithmetic.

- Deficits in working memory:

Working memory is defined as the system which actively holds information in the mind to do verbal and nonverbal operations such as reasoning and comprehension. Working memory tasks are those that require the goal-oriented active monitoring or manipulation of information in the face of interfering processes and distractions. The cognitive processes involved include

the executive and attention control of short-term memory which provide for the interim integration, processing, disposal, and retrieval of information.

- Deficits in processing speed (especially in counting rate):

Processing Speed is one of the measures of cognitive efficiency. It includes the ability to automatically and fluently perform relatively easy or over-learned cognitive tasks, especially when high mental efficiency is required. That is, for simple tasks requiring attention and focused concentration. It relates to the ability to process information automatically, without intentional thinking through. A child with processing speed needs has difficulty in performing simple cognitive tasks fluently and automatically, especially when mental efficiency in focusing concentration is required. Students with processing speed needs may take more time to: recognize simple visual patterns and in visual scanning tasks; take tests that require simple decision making; perform basic arithmetic calculations and in manipulating numbers, since these operations are not automatic for them; perform reasoning tasks under time pressure; make decisions that require understanding of the material presented; read silently for comprehension; copy words or sentences correctly or to formulate and write passages.

Other authors include problems with the visual-spatial motor organization in the group of deficits in mathematical learning. Dr. Kate Garnett, director of the masters programs in Learning Disorders at Hunter College in USA, underline the importance of visual-spatial aspects of math regarding math learning disorders. In the *Division for Learning Disabilities Journal of CEC* (1998), she writes:

Disturbances in visual-spatial-motor organization may result in weak or lacking understanding of concepts, very poor "number sense," specific difficulty with pictorial representations and/or poorly controlled handwriting and confused arrangements of numerals and signs on the page. Students with profoundly impaired conceptual understanding often have substantial perceptual-motor deficits and are presumed to have right hemisphere dysfunction. (p.19)

This small subgroup may well require a very heavy emphasis on precise and clear verbal descriptions. They seem to benefit from substituting verbal constructions for the intuitive/spatial/relational understanding they lack. Pictorial examples or diagrammatic explanations can thoroughly confuse them, so these should not be used when trying to teach

or clarify concepts. In fact, this subgroup is specifically in need of remediation in the area of picture interpretation, diagram and graph reading, and nonverbal social cues. To develop an understanding of math concepts, it may be useful to make repeated use of concrete teaching materials (e.g., Stern blocks, Cuisenaire rods), with conscientious attention to developing stable verbal renditions of each quantity (e.g., 5), relationship (e.g., 5 is less than 7), and action (e.g., $5+2=7$). Since understanding visual relationships and organization is difficult for these students, it is important to anchor verbal constructions in repeated experiences with structured materials that can be felt, seen, and moved around as they are talked about.

It is important to mention that the classification of mathematical deficits presented above is one of the many classifications that have been proposed by the researchers. There is no consensus about the different kinds of deficits in the process of mathematical learning.

Now, regarding the causes of mathematics disorders, it is still a polemic topic, and there are different perspectives from which the causes of the disability can be analyzed: Neurobiological causes, Information processing causes and Information processing causes.

Neurological and other biological causes:

The neurological approach argues that the basis of the difficulties in mathematics is a defect roughly notorious at a neurological level. In the first investigations about this subject, "dyscalculia" was considered a derivation of "acalculia" or blindness for numbers. Some years after, Stadelmann and Lewanolowsky proposed, based on their research, the left occipital region as the "center of arithmetic abilities".

Following this neurological perspective Hecaen, Houillier and Angelerques proposed a tripartite organization based on neuropsychological mechanisms underlying each type:

- Type 1. Acalculia: resulting from alexia and agraphia for numbers in which the patient is unable to write or read the number needed to perform the calculation.
- Type 2. Acalculia of spatial type: associated with impaired spatial organization of numbers such as incorrect alignment of the digits.
- Type 3. Anarithmic: it consists in an inability to perform arithmetic procedures despite having intact visual-spatial skills and abilities to read and write numbers.

During the decades of the 60s and 70s, researchers started to relate the mathematical disabilities to linguistic disorders. For example, Cohn (1968) proposed that the mathematical disabilities were part of a more general linguistic dysfunction caused by a lack of coordination of several complex neurological systems.

In 1974, Kosc developed a classification which integrated six subtypes of dyscalculia, which may occur in isolation or in combination:

- Verbal dyscalculia: difficulties in naming mathematical quantities, numbers, terms, symbols and relationships.
- Practognostic dyscalculia: difficulty to enumerate, compare and manipulate objects mathematically.
- Lexical dyscalculia: difficulty in reading mathematical symbols.
- Graphic dyscalculia: difficulty in writing mathematical symbols.
- Ideognostical dyscalculia: difficulty making mental operations and in understanding mathematical concepts.
- Operational dyscalculia: difficulty in performing operations and numerical calculations.

Finally, it is important to mention that one of the most investigated aspects from the neurological perspective has been the brain lateralization in mathematics disorders. Many researchers believe that the performance of the right hemisphere in learning mathematics is essential since it is specialized in organization and visual-spatial integration, both essential for proper performance in arithmetic.

As we can see, the neurological perspective of the causes of the mathematical learning disabilities has been criticized for conducting researches which are not based on sound theory on mathematical competence, and for the lack of experimental controls and methodological rigor in order to get serious conclusions. Some professionals even consider this perspective very unfruitful because it provides no information on the number of faulty cognitive processes that constitute the immediate causes of the poor performance of the persons with mathematics disorders.

Information processing causes:

According to this perspective, if we know the mental processes that are used to perform an operation or intellectual structures that the student must carry it out, we can better understand where and why they make mistakes. The aim is to understand and explain what the learner does.

To complement this perspective of information processing, there is the so-called theories of parallel processing (PDP) which claim that it is necessary to make an analogy between the computer and the brain's neural connections (Rumelhart, McClelland and the PDP Group, 1992) . These theories state that the information processing is performed by a large number of interacting units, since they are connected form a network characterized by the general level of activation produced by the input signal and the connection strength between each of the units.

Another very important feature of this perspective, called also connectionist, is the idea that besides the input and output units of information that the brain has to connect with the environment, there is an existence of hidden units that are carrying the weight of the cognitive work of the brain system.

As stated Garcia Madruga (1992) connectionist theories are fully adapted to the explanation of evolutionary phenomena and therefore from this approach, learning is the establishment of new connection of networks between the different interactive units.

Environmental causes:

From this perspective, the stimulation that a child receives in the early stages of his life may be decisive to prevent or encourage the development of disorders such as mathematical learning difficulties. One of the main representatives of this perspective is the theory of Karmiloff-Smith which makes a restatement of the theory of Fodor (1986) and argues that the mind possesses innate architecture specification called "modules". These modules correspond to some genetically specified performances.

So according to this point of view, to understand the learning difficulties in mathematics it is essential to consider the context in which they occur; and therefore to identify children with

mathematical learning disabilities, it is necessary to examine the mathematics instruction or teaching in the classroom. It is important to develop systematic methods of analysis and evaluation of classroom instruction and review the context of possible causes of the low performance of the child.

One of the methods used in this perspective to find the causes of the disabilities is based on the concept of the zone of proximal development developed by Vygotsky (1979). According to this theory, the amount of aid that the student needs is an estimate of the effectiveness of learning in that domain. The assessor continues to help the student until he is able to solve problems independently.

Under the environmental perspective, we cannot focus only in biological or cognitive cause since feelings, beliefs, social surroundings like family, school, etc, also determine the performance of a child and therefore we could say that mathematics disorders are socially constructed.

As we can see, researchers are still unsure of the exact causes of mathematics disorders. Multiple causative factors have been proposed including psychological, neurological, genetic, and social factors. Some researchers propose that the disorder is the product of a mixture between these different factors. There is still much research that needs to be done to fully understand the causes and how to prevent mathematical disorders.

Chapter 3: The development of Psychometrics.

This chapter presents an overview of psychometrics, including its history alongside the main theories of tests. It also describes briefly the standards for the educational and psychological testing developed by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME).

3.1 A historical review of Psychometrics.

The origin of psychometrics can be placed at the end of the nineteenth century with Francis Galton (1883) who is considered to be the precursor of this branch of psychology. Thereafter, psychometrics will be developed mainly through psychophysical studies. These studies led to the development of models that allowed assigning numerical values to stimuli and psychological characteristics, and developing the different theories of tests.

We can identify 3 factors as decisive in the development of tests:

- The opening of the anthropometric Galton Laboratory in London in 1904.
- The development of the Pearson correlation measurement.
- Spearman's interpretation about Pearson correlation, explaining that the correlation between two variables indicates that both have a common factor.

It is important to mention also that the use of tests as instruments anticipated their own theoretical foundation. The first use of these instruments is related to the sensorimotor tests used by Galton (1822-1911) in his anthropometric laboratory in Kensington. Galton was also a pioneer in using statistical techniques to analyze the data from his tests. Years later, Pearson continued developing this kind of statistical techniques.

The psychologist, James McKeen Cattell (1860-1944) was the first to use the term "mental test", but his tests as well as those of Dalton were most sensory tests and the analysis of the data obtained through them left clear that there was no correlation between these scales and the intellectual level of the subjects to whom the scales were administered. It is Alfred Binet who in 1905 made a radical turn in the philosophy of the tests, by introducing in his scale more cognitive tasks designed to assess aspects such as judgment and intelligence. In 1916, at Stanford University the psychologist Lewis Terman released a revised examination of the Binet-Simon test which became known as the "Stanford-Binet test". The Binet test was the first to use the Intelligence Quotient ratio (IQ) to express the scores of the

subjects. The concept of the IQ was originally from Stern, who in 1911 proposed dividing mental age (MA) between the chronological age (CA) by multiplying by a hundred to avoid decimals: $CI = (EM / EC) \times 100$.

The next step in the historical development of the tests will be marked by the emergence of collective intelligence tests, prompted by the need for the U.S. Army in 1917 to select and classify the soldiers who were going to take part in the First World War. A committee led by Yerkes designed, based especially on Otis unpublished test, the now famous Alpha and Beta tests. The Alpha test is for the general population and the Beta is for the use with inmates who are illiterate or not fluent in English. Both of these tests are still in use.

The emergence of the classical test batteries as we know them today, were developed around the 30s and 40s. One example of the tests developed in these decades is the Primary Mental Abilities test constructed by Thurstone. The different models developed during these years gave rise to numerous batteries of tests (PMA, DAT, GATB, TEA, etc) which are commonly used also nowadays (Rust,2009).

3.2 Psychometrics' main theories.

In the psychometrics' field there have been developed a number of different measurement theories. These include classical test theory (CTT) and item response theory (IRT). An approach which seems to be mathematically different to IRT is represented by the Rasch model for measurement. The development of the Rasch model, and the broader class of models to which it belongs, was explicitly founded on the requirements of measurements in the physical sciences.

Researchers have also developed methods for working with large matrices of correlations and covariances. Techniques in this tradition include: factor analysis, multidimensional scaling, and data clustering. All these multivariate descriptive methods try to distill large amounts of data into simpler structures. More recently, structural equation modeling and path analysis represent more sophisticated approaches to working with large covariance matrices. These methods allow statistically sophisticated models to be fitted to data and tested to determine if they are adequate fits.

Here below is presented a brief description of the two main theories, the classical test theory and the items response theory:

Classical test theory

It is a psychometric theory that predicts outcomes of psychological testing such as the difficulty of items or the ability of test-takers. The aim of classical test theory is to understand and improve the reliability of psychological tests.

Classical test theory as we know it today was codified by Novick (1966) and described in classic texts such as Lord & Novick 1968) and Allen (1979).

Classical test theory assumes that each person has a true score, T , that would be obtained if there were no errors in measurement. A person's true score is defined as the expected number-correct score over an infinite number of independent administrations of the test.

Unfortunately, test users never observe a person's true score, only an observed score, X . It is assumed that observed score = true score plus some error.

Items response theory

Item response theory (IRT) also known as strong true score theory, is a paradigm for the design, analysis, and scoring of tests, questionnaires, and similar instruments measuring abilities, attitudes and other variables. It is based on the application of related mathematical models to testing data.

The name *item response theory* corresponds to the focus of the theory on the item, as opposed to the test-level focus of classical test theory, by modeling the response of an examinee on a given ability to each item in the test. The term item is used because many test questions are not really questions; they might be multiple choice questions that have incorrect and correct responses, also commonly statements on questionnaires that allow respondents to indicate level of agreement (a rating or Likert scale), or patient symptoms scored as present/absent.

IRT is based on the idea that the probability of a correct/keyed response to an item is a mathematical function of person and item parameters. The person parameter is known as latent trait or ability and it may represent a person's intelligence or the strength of an attitude.

The purpose of IRT is to provide a framework for evaluating how well assessments work, and how well individual items on assessments work. The most common application of IRT is in

education, where researchers use it for developing and refining exams, maintaining banks of items for exams, and equating for the difficulties of successive versions of exams.

3.3 Standards for Educational and Psychological Testing.

The Standards for Educational and Psychological Testing is a set of testing standards developed jointly by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME). The collaboration of the three associations has been formalized in a cooperative agreement that creates a management structure and sets procedures for maintaining and revising the Standards.

The *Standards* is written for the professional and for the educated person and addresses professional and technical issues of test development and use in education, psychology and employment. The *Standards* are currently under review once again. A revised version is expected to be published sometime after 2012.

The purpose of publishing the Standards is to provide criteria for the evaluation of tests, testing practices, and the effects of test use. Although the evaluation of the appropriateness of a test or testing application should depend heavily on professional judgment, the Standards provide a frame of reference to assure that relevant issues are addressed.

It is hoped that all professional test developers, sponsors, publishers, and users will adopt the Standards and encourage others to do so. Overall, the Standards advocate that, within feasible limits, the relevant technical information should be made available so that those involved in policy debate may be fully informed.

The *Standards* are organized into three major sections. The document begins with a series of chapters devoted to the test development process, which focus primarily on the responsibilities of test developers. The *Standards* continue with chapters that address specific uses and applications of the standards, which focus primarily on the responsibilities of test users. Each chapter within the major sections begins with an introductory text, providing an explanatory background for the standards that follow.

A summary of the Standards is presented below:

Part I of the Standards: Test Construction, Evaluation, and Documentation

1. Standards for Validity

The standards refer to validity as the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is the most fundamental consideration in developing and evaluating tests. Professional judgment guides decisions regarding the specific forms of evidence that best support the intended interpretation and use.

2. Reliability and Errors of Measurement

Reliability refers to the consistency of measurements when the testing procedure is repeated on a population of individuals or groups. However, no single examinee is completely consistent, and in some instances, because of subjectivity in the scoring process, an individual's obtained score and the average score of a group will always reflect at least a small amount of measurement error. Information about measurement error is essential to the proper evaluation and use of a test instrument.

3. Test Development and Revision

Test development is the process of producing a measure of some aspect of an individual's knowledge, skill, ability, interests, attitudes, or other characteristics by developing items and combining them to form a test, according to a specified plan. Test development also includes specifying conditions for administering the test, determining procedures for scoring the test performance, and reporting the scores to test users.

4. Scaling, Norming, and Scale Comparison.

Scale scores may aid in the interpretation of scores by indicating how a given score

compares to those of other test takers, by enhancing the comparability of scores obtained using different forms of a test or in other ways.

5. Test Administration, Scoring, and Reporting

The standards explain that the usefulness and interpretability of test scores require that the directions to examinees, testing conditions, and scoring procedures follow the same detailed procedures. When these steps are taken, the test is said to be standardized, without such, the accuracy and comparability of score interpretations would be reduced.

6. Supporting Documentation for Tests

The standards describe that test documents need to include enough information to allow test users and reviewers to determine the appropriateness of the test for its intended purposes.

A test's documentation typically specifies the nature of the test; its intended use; the process involved in the test's development; technical information related to scoring, interpretation, and evidence of validity and reliability; scaling and norming if appropriate to the instrument; and guidelines for test administration and interpretation.

Part II of the Standards: Fairness in Testing

1. Standards on Fairness and Bias

The standards focus on the aspects of fairness and testing that are customarily the responsibility of those who make use and interpret tests, which are characterized by some level of professional and technical consensus. It does not examine the very broad issues related to regulations, statutes, and case law that govern test use and the remedies for harmful practice.

These standards describe fairness in the following four principle ways in which the term

fairness should be used: Fairness as a Lack of Bias; Fairness as Equitable Treatment in the Testing Process; Fairness as Equality in Outcomes of Testing; and Fairness as Opportunity to Learn.

The standards concerning bias describe the term bias as construct-irrelevant components that result in systematically lower or higher scores for identifiable groups of examinees. Likewise, two main sources of bias are identified: Content-Related Sources of Bias and Response-Related Sources of Bias.

2. The Rights and Responsibilities of Test Takers

Fairness issues unique to the interests of the individual test taker, which reflects widely accepted principles in the field of measurement are also addressed. The responsibilities of test takers concerning test security, their access to test results, and their rights when irregularities in their testing are claimed are discussed.

3. Testing Individuals of Diverse Linguistic Backgrounds

Any test that employs language is, in part, measuring a test taker's language skills. It is important to consider language background in developing, selecting, administering, and interpreting test performance.

4. Testing Individuals with Disabilities

Although the Standards focus on technical and professional issues regarding the testing of individuals with disabilities, test developers and users are also encouraged to become familiar with federal, state, and local laws, and court and administrative rulings that regulate the testing and assessment of individuals with disabilities.

However, the Standards do address issues regarding appropriate accommodations when testing individuals with disabilities, strategies of test modification, using modifications in different testing contexts, and reporting scores on modified tests.

Part III of the Standards: Testing Applications

1. Standards Involving General Responsibilities of Test Users

Test users are referred to as the group of professionals who actively participate in the interpretation and use of test results. In selecting a test and interpreting a test score, the test user is expected to have a clear understanding of the purposes of the testing and its possible consequences.

2. Psychological Testing and Assessment

This chapter of the standards addresses issues important to professionals who use psychological testing with their clients. The topics covered include test selection and administration, test interpretation, collateral information used in psychological testing, types of tests, and purposes of testing.

The types of psychological tests described include: Cognitive and Neuropsychological Testing; Social, Adaptive, and Problem Behavior Testing; Family and Couples Testing; Personality Testing; Vocational Testing; Interest Inventories; Work Values Inventories; and Measures of Career Development, Maturity, and Indecision. The purposes of psychological testing are outlined as Testing for Diagnosis; Testing for Intervention Planning and Outcome Evaluation; Testing for Judicial and Governmental decisions; and Testing for Personal Awareness, Growth, and Action.

3. Educational Testing and Assessment

Within this context, the standards are concerned with testing in formal educational settings from kindergarten through post-graduate training. In this section, three broad areas of educational testing are considered, which encompass various given purposes of educational testing. They are: (a) routine school, district, state, or other system-wide testing programs, (b) testing for selection in higher education, and (c) individualized and special needs testing.

4. Testing in Employment and Credentialing

The standards describe employment testing as being carried out by organizations for purposes of employee selection, promotion or placement. These three purposes all focus on the prediction of future job behaviors, with the goal of influencing organizational outcomes, such as efficiency, growth, productivity, and employee motivation and satisfaction. The Standards address testing for licensure and certification, with a focus on the applicant's current skill or competency in a specified domain.

5. Testing in Program Evaluation and Public Policy

Tests are also widely used in program evaluation, which assesses the needs, implementation, effectiveness, and value of a program, and in public policy decision making. Test results can be utilized as a source of evidence for the initiation, continuation, modification, termination, or expansion of various programs and policies. The Standards recognize that it is important to evaluate any proposed test in terms of its relevance to the goals of the program or policy and/or to the particular question its use will address.

Chapter 4: The Utrecht Early Mathematical Competence Test.

This chapter describes the main features of the Utrecht Early Mathematical Competence Test. Before the first subtitle of this section, a brief introduction of the EMCT will be presented below:

The Utrecht Early Mathematical Competence test was developed by Hans van Luit; Bernadette van de Rijt & Albèr H. Pennings in 1994, Netherland. The authors of the test are staff members of the Department of Special Education at the Utrecht University.

This scale represents a task-orientated test which aims to measure the level of early mathematical competence. The test has been developed for Kindergarten 1st and 2nd year and grade 1 of the primary school. The test is not tied to a certain mathematical course nor to a mathematical method. The test consists of two parallel versions (versions A and B) of 40 items each. Each scale or version consists of eight parts, the tasks are spread over these parts in groups of five.

The EMTC should be examined individually. With the help of Version A and Version B the teacher or another user of the test, is able to keep track of the development of early mathematical competence of a child or a group of children. By comparing the performance of a child with those of children in a norm group, the level of early mathematical competence can be determined.

A possibility for application of both versions of the test is the verification of a non-expected test score. The results of a test administration by a toddler can also be determined by the circumstances in which the administration has taken place and can therefore be unexpectedly high or low. In order to see if this is the case, the parallel version can be examined after a few days. If certain circumstances played no role in the first administration, then the second administration will give a similar score. If circumstances indeed played a role in the first administration, then the second administration will give a result that is perhaps more in the line of expectation.

In addition to the determination of the level of early mathematical competence, the teacher is able, also with the help of both versions of the test, to ascertain if the children make progress under the influence of a mathematical course or mathematical method. By examining Version A and Version B at respectively the beginning and the end of the education program, the

teacher is able to ascertain to what extent the child has reached a higher level of early mathematical competence.

4.1 Structure of the test.

4.1.1 Concepts of the test.

As it is mentioned above, each version of the test is constituted by eight specific concepts. Through these 8 concepts, the test or scale is divided in 8 sub-scales with 5 items in each of them.

The components of the EMTC are: Concepts of comparison, Classification, Correspondence, Seriation, Using counting words, Structured counting, Resultative counting, and General knowledge of numbers.

Concepts of Comparison:

The comparison of quantitative or qualitative characteristics of objects. With this component it is ascertained if children master the concepts which are frequently used in comparisons, especially in mathematical education; concepts like 'the most', 'the least', 'higher', 'lower'

Classification:

The grouping of objects in class or subclass based on criteria, as clearly as possible. With the classification task it is ascertained if children are able, on the basis of similarity or differences, to distinguish between objects and group them.

Correspondence:

The comparison of amounts by making an one-to-one-relation. With this component it is ascertained if children are able to make a one-one-relation between different objects. For

example: are there as many chickens as there are eggs? At the same it is ascertained if children understand that six blocks is the same quantity as six points on a dice.

Seriation:

The ranking of objects in class or subclass based on criteria, as clearly as possible. With this component it is ascertained if children are capable of recognizing the correct rank order. Terms used in the tasks are: from high to low, from more to less, from thin to thick, from small to broad. Next to this, children have to make a series of their own by means of drawing lines from, for example, a big rabbit to a big carrot and from a small rabbit to a small carrot.

Using counting words:

Counting forwards, counting backwards and counting on as well as using the cardinal and ordinal number. With the tasks of this component the acoustic counting is examined and next to this it is ascertained if children know how to make use of cardinal and ordinal numbers till twenty.

Structured counting:

Synchronous counting, shortened counting from the dice structure. With this component it is ascertained, by making use of objects (blocks among other things), if children master synchronous counting of quantities. At this component the children are allowed to point to the objects with their fingers while counting. Next to this it is ascertained with this component if certain dice structures are recognized immediately.

Resultative counting:

Counting structured and unstructured quantities as well as counting hidden quantities.

With this component it is ascertained if children are capable of determining the total number of objects from both structured and unstructured collections. While counting they are not allowed to use their fingers for pointing to the objects in the collection.

General knowledge of numbers:

Being able to use knowledge of the number system in simple problem situation. With this component it is ascertained if children are able to use numbers under twenty in simple daily problem situations.

4.1.2 Items and instructions.

The tasks described in the items in Version A and B are developed by the test designers themselves on the basis of a study of psychological research concerning the development of early mathematical competence as well as mathematical courses and mathematical methodology. The tasks in Version A and B are derived from one item bank. Both versions measure the same. The versions mainly consist of verbal presented tasks and the child is able to point out the right answer. Next to this there are tasks where the child only has to say the answer, and finally there are a small number of manipulation tasks. A distinction can be made in tasks where the child works with material and tasks where it makes certain actions itself. Each of the tasks is scored 'correct' or 'incorrect'.

Since in Ecuador is the Scale A the one that is used, and in view of the fact that for this research the reliability of the Spanish version of the scale A was measured, the following items that will be described correspond to the Scale A, and the items of scale B will not be described here.

I) Concepts of comparison:

Material: UEMCT cards (Appendix A)

Task A1. Here you see mushrooms. Point out the mushroom which is higher than this flower.

[The teacher points out the flower in the square at the left top of the page].

Task A2. Here you see men. Point out the man who is thicker than this man. [The teacher points out the man in the square at the left top of the page].

Task A3. Here you see buildings. Point out the lowest building.

Task A4. Here you see indians. Point out the Indian who has less feathers than this Indian with bow and arrow. [The teacher points out the Indian in the square at the left top of the page].

Task A5. Here you see boxes with marbles. Point out the box with the fewest marbles.

II) Classification

Material: UEMCT cards.

Task A6. Look at these pictures. Point out the picture with something that cannot fly.

Task A7. Look at these squares. Point out the square with five blocks but without triangles.

Task A8. Look at these pictures. Point out all the grey circles.

Task A9. Here you see people. Point out all the people who have a bag, but no glasses.

Task A10. Here you see an apple with a stalk without a leave and a little worm coming out of the apple. [The teacher points out the apple in the square at the left top of the page.] Point out all the apples which are exactly the same.

III) Correspondence

Material: The material consists of blocks for task 11 and 12, two work papers and a pencil for task 13 and 14.

Task A11. [The teacher gives 10 blocks to the child.] You have thrown four with a dice. [The teacher shows the four-structure of a dice.] Can you lay down the same amount of blocks?

Task A12. [The teacher gives the child 15 blocks.] I have thrown with two dice these points. Can you lay down the same amount of blocks? [The teacher shows the picture of two dices with a 5-structure and a 6-structure.]

Task A13. [The teacher gives the child a work paper and a pencil.] Here you see candle-holders and candles. In each candle-holder candles fit in. Can you draw lines from the candles to the candle-holder in which they fit in?

Task A14. [The teacher gives the child a work paper and a pencil.] Here you see three pictures of chickens and eggs. Can you find the picture where each chicken has laid one egg? You may draw lines.

Task A15. Here you see fifteen balloons. [The teacher points out the balloons in the square on the left top of the page.] Point out the square where there are as many dots as there are balloons.

IV) Seriation

Material: The material consists of a work paper with a pencil for task 19 and then UEMCT cards for the rests of the cards.

Task A16. Here you see squares with apples in it. Point out the square where the apples are ranked from big to small.

Task A17. Here you see squares with rocks. Point out the square where the rocks are ranked from thin to thick.

Task A18. Here you see squares with marbles. Point out the square where the marbles are ranked from small and brightly colored to big and dark colored.

Task A19. [The teacher gives the child a work paper and a pencil.] Here you see dogs. Each dog is going to get a stick. A big dog is going to get a big stick and a small dog is going to get a small stick. Can you draw lines form the dogs to the sticks they are going to get?

Task A20. Here you see slices of bread in a row from many to few slices of bread. [The teacher Points out the row of slices of bread at the bottom of the page] These slices of bread fit somewhere in the row, [The teacher points out the slices of bread in the square at the left top of the page] point out where in this row these slices of bread fit in.

V) Using counting words

Material: UEMCT cards

Task A21. Count up to twenty.

Task A22. [The teacher shows the picture to the child.] Point out the square with seven dots.

Task A23. Count on from nine to fifteen: six, seven, eight

Task A24. [The teacher shows the picture to the child.] Point out the eighteenth flower.

Task A25. Count up to 14 with skipping one each time: two, four, six

VI) Structured counting (Counting synchronously and shortened counting)

Material: The material consists of blocks for task 26, 27, 28 and 30.

Task A26. [The teacher lays down 16 blocks on the table in four rows of four blocks each and with a small distance between the blocks.] Point out the blocks and count them.

Task A27. [The teacher lays down nine blocks on the table in a circle with a small distance between the blocks.] Count these blocks. [The child is allowed to point out the blocks or to push them aside while counting.]

Task A28. [The teacher lays down 20 blocks on the table in a heap with a small distance between the blocks.] Count these blocks. [The child is allowed to point out the blocks or to push them aside while counting.]

Task A29. I'll show you a picture you have to take a good look at for a short while. [The teacher shows the picture to the child for 2 seconds (count: twenty-one, twenty-two) and removes the picture.] How many dots are there on the dice?

Task A30. [The teacher lays down 17 blocks on the table in a row with a small distance between the blocks.] Here you see seventeen blocks. Point out these blocks and count backwards.

VII) Resultative counting

Material: The material consists of blocks for all five tasks.

Task A31. [The teacher gives the child 15 blocks.] Lay down a row of eleven blocks.

Task A32. [The teacher lays down 20 blocks on the table in a row with a small distance between the blocks.] How many blocks are lying here? [The child is not allowed to point out the blocks.]

Task A33. [The teacher lays down 15 blocks on the table in three rows within each row five blocks and with a small distance between the blocks. How many blocks are lying here? [The child is not allowed to point out the blocks.]

Task A34. [The teacher lays down 19 blocks in a heap on the table with a small distance between the blocks.] How many blocks are lying here? [The child is not allowed to point out the blocks.]

Task A35. [The teacher lays down 5 blocks on the table.] Here are five blocks. I push them under my hand. [The teacher pushes the blocks under his hand. Then he pushes 7 blocks, which he shows to the child, also under his hand.] I add seven blocks. How many blocks are there under my hand?

VIII) General knowledge of numbers

Material: UEMCT Cards.

Task A36. Here you see two boxes. [The teacher points out the boxes on the picture.] In the black box are nine candies. In the white box are thirteen candies. In which box are the most candies?

Task A37. [The teacher points out the picture with the 9 marbles.] You have nine marbles. You lose three marbles. How many marbles do you have left? Point out the square with the right number of marbles. [The teacher points out the row on the bottom of the page with the pictures.]

Task A38. [The teacher points out the picture with the 8 chickens.] A farmer owns eight chickens. He buys two chickens. [The teacher points out the picture with two chickens.] How many chickens do the farmer owns now? Point out the square with the right number of chickens. [The teacher points out the row of pictures on the bottom of the page.]

Task A39. Here you see a building. In the building there are windows. [The teacher points out the windows in the building, not one by one.] There are also trees standing in front of the building. Can you count how many windows the building has?

Task A40. [The teacher points out the picture of a dice.] This is a dice. You have thrown two dices. [The teacher points out the two dices.] Look how many dots you have thrown and point out where the pawn should be standing.

4.1.3 Norms and Interpretations of the results.

The raw total score of the test is the total number of tasks in Version A or Version B (depending on which version has been administered) which are answered correctly by the child. The test score of the child is transformed in a competence score. This competence score indicates how the child masters early mathematical competence. A relatively low score indicates a low early mathematical competence, a relatively high score indicates a high early mathematical competence. The competence score doesn't give enough information by itself.

Its meaning must be derived from making comparisons with scores of children in the same class or age group.

To give the test user a basis for comparison, the tasks from Version A or B were administered in the period of January-February by 823 children from 25 primary schools in Nederland. On the basis of the results it has been possible to develop some general norms. Since it is not always clear how long the child has attended education, norms have been developed for six half-year age groups between 4 years and 7 months up to and including 6 years and 12 months.

In table 1, limits for the age groups (I - VI) are presented:

Table 1: Age groups.

Limits of the six age groups	
Group	Limits
I	4.07-4.12
II	5.01-5.06
III	5.07-5.12
IV	6.01-6.06
V	6.07-6.12

The steps to determine the level of early mathematical skills of a child according to the test scores are:

- Examine the child with one of both versions.
- Evaluate the answers with the help of a score key-answer sheet for Version A or B; after this, the number of correct answers given is determined. This number indicates the test score of the child.

- With help of the information given in the Scale scoring table (appendix B), one looks for the competence score belonging to the test score.
- After having determined the competence score and identifying in which age level group the child is located, the test user looks up to the adequate norm table. In this table he/she looks up the competence score of the child. At the beginning of the table he/she will find on the same line as the competence score, the level of the child.

For Version A, the norm table would be the following:

Table 2: Norm table for Version A.

	I	II	III	IV	V	VI
Level	N=178	N=155	N=128	N=152	N=111	N=99
A	more than 48	more than 63	more than 65	more than 77	more than 82	more than 83
B	41-48	55-63	59-65	70-77	75-82	76-83
C	32-40	46-54	50-58	60-69	70-74	70-75
D	24-31	38-45	43-49	49-59	62-69	64-69
E	less than 24	less than 38	less than 43	less than 49	less than 62	less than 64

As a result of the application of the norm table, the test score of the child will be transformed in an indicator of the level of the child. The test score of Version A or B is first transformed into a competence score. Then the test user determines the level of the competence score. Therefore the competence score is compared with the scores of children in the same norm group. The levels have been determined as follows:

Level A: good to very good (comparable with around 25% of the highest scoring children in the norm group).

Level B: ample to good (comparable with around 25% of the children in the norm group scoring just above the average)

Level C: moderate to ample (comparable with around 25% of the children in the norm group scoring just below the average).

Level D: weak to moderate (comparable with around 15% of the children in the norm group scoring ample below the average).

Level E: very weak to weak (comparable with around 10% of the lowest scoring children in the norm group).

When the level of the child has been determined, the test user has a complete view of how the competence score of the child lays in comparison with the competence scores of the norm group children. Then his or her place in the rank order of the norm group is known. Children who score on level A for example have an achievement which is comparable with the achievement of only the top 25% of the norm group. Children who achieve level D, are outdone by at least 75% and at the most 90% of the norm group.

4. 2 Background of the test.

The Utrecht Early Mathematical Competence Test was constructed based on the ideas of the Swiss developmental psychologist Jean Piaget, the Realistic Mathematics Education movement (RME) and the main concepts in the development of children's counting generated by Gelman and Gallistel.

An outline of the background knowledge used by the authors of this test is presented below:

Piaget's mathematical perspective:

In his research about the development of the fundamental elements of early mathematical competence, Jean Piaget stated that conservation of number is the minimum criterion for early mathematical competence.

Early mathematical competence determined by conservation-of-number tasks, is based on a complex synthesis of classifying objects operating with the one-to-one-relation principle (correspondence) and ranking objects (seriation). Conservation-of-number requires also insight in the cardinal and ordinal component of the number.

A cardinal number refers to the total number of objects in a collection (five balloons) while an ordinal number represents the position of one object with respect to the other objects in the collection (the fifth balloon).

Realistic Mathematics Education movement (RME):

Realistic Mathematics Education, or RME, is a Dutch movement that was born to reform the teaching and learning of mathematics. The roots of the Dutch reform movement go back to the early seventies when the first ideas for RME were conceptualized. It was a reaction to both the American "New Math" movement that was probable to flood Netherlands in those days, and to the then prevailing Dutch approach to mathematics education, which often is labeled as "mechanistic mathematics education".

The development of what is now known as RME started almost thirty years ago. The foundations for it were settled by the Dutch researcher Hans Freudenthal and his colleagues at the former IOWO, which is the oldest predecessor of the Freudenthal Institute.

According to Freudenthal, mathematics must be connected to reality, stay close to children and be relevant to society, in order to be of human value. Instead of seeing mathematics as subject matter that has to be transmitted, mathematics should be seen as a human activity. Education should give students the guided opportunity to "re-invent" mathematics by doing it. This means that in mathematics education, the focal point should not be on mathematics as a closed system but on the activity, on the process of mathematization (Freudenthal, 1983).

Later on, Treffers (1978, 1987) formulated the idea of two types of mathematization explicitly in an educational context and distinguished "horizontal" and "vertical" mathematization. In broad terms, these two types can be understood as follows:

In horizontal mathematization, the students come up with mathematical tools which can help to organize and solve a problem located in a real-life situation.

Vertical mathematization is the process of reorganization within the mathematical system itself, like, for instance, finding shortcuts and discovering connections between concepts and strategies and then applying these discoveries.

In short, one could say that horizontal mathematization involves going from the world of life into the world of symbols, while vertical mathematization means moving within the world of symbols.

Children's counting development:

The American psychologists Fuson (1988), Gelman and Gallistel (1978) and the Russian psychologist Davydov have investigated the development of Counting (Van Parreren, 1989). Much consensus is found between the opinions of researchers mentioned above concerning the findings with regard to the phases in the development in counting and the age attending these phases:

Phase 1. Acoustic counting.

At the age of about three, children begin with acoustic counting; the counting is nothing more than reciting a poem or a song.

Phase 2. Asynchronous counting.

At the age of about four the asynchronous counting manifests itself. Children use the numbers in the right order, but they are not able yet to point to one object while enumerating one number. Frequently they miss an object or point to the same object twice. Counting and pointing to objects at the same time is not yet possible. When it is possible, they are able to count synchronously.

Phase 3. Arranging objects while counting.

When an amount of unarranged objects has to be counted, children start to arrange the objects while counting. For instance, they push the counted objects aside. Children at the age of four and a half master this arranged counting.

Phase 4. Resultative counting.

At the age of five, children reach the phase of resultative counting. This means that they are aware of the fact that counting must begin with number one, that every object must be counted once, and that the last number mentioned gives the total amount of objects. Important in this phase is the fact that the children discover the one-to-one-relation between object and number.

Phase 5. Shortened counting.

After resultative counting, children learn another strategy for counting that is shortened counting. In a number of objects the children recognize the representation of, for example, the five on the dice, and they count from this number on. Children at the age of five and a half years to six should be able to deal with shortened counting. It has been found that children at the age of five and six years old also are able to solve conservation-in-number tasks and correspondence tasks by making use of counting.

As we can see the UEMCT is based on a rich theoretical framework, which includes various theories on the understanding of mathematics in children and on the process of cognitive development. In the next and last chapter of this dissertation, we will see the results regarding the reliability of the Spanish version of the UEMCT within an Ecuadorian kindergarten context.

Chapter 5: Statistical estimations, analysis and conclusions.

In this last chapter the statistical approach applied for this study is described, including the correspondent explanations of the different estimations' processes and other implications of the statistical analysis. The last section of this chapter presents a report of the findings and relevant results obtained in this study and the interpretations, conclusions and recommendations inferred from these results.

5. 1. Reliability.

In statistics, reliability refers to the consistency of a measure. A test is considered reliable if we get the same or a very similar result repeatedly. For example, if a test is designed to measure a trait (such as introversion), then each time the test is administered to a subject, the results should be approximately the same. Unfortunately, it is impossible to calculate reliability exactly, but it can be estimated in a number of different ways.

Before describing the ways to estimate the reliability of a test, it is necessary to mention that the reliability of a psychometric test is presented as a coefficient. The reliability coefficient range from 0 to 1.0. A coefficient of 0 means no reliability and 1.0 means perfect reliability. Since all tests have some error, reliability coefficients never reach 1.0. Generally, if the reliability of a standardized test is above .80, it is said to have very good reliability; if it is below .50, it would not be considered a very reliable test. The different estimates of reliability are:

- Inter-rater reliability: is the variation in measurements when taken by different persons but with the same method or instruments.
- Test-retest reliability: is the variation in measurements taken by a single person or instrument on the same item and under the same conditions. This includes intra-rater reliability. In the test-retest method, reliability is estimated as the Pearson product-moment correlation coefficient between two administrations of the same measure.

- Inter-method reliability: is the variation in measurements of the same target when taken by a different methods or instruments, but with the same person, or when inter-rater reliability can be ruled out. When dealing with forms, it may be termed parallel-forms reliability.
- Internal consistency reliability: assesses the consistency of results across items within a test. The most common internal consistency measure is Cronbach's alpha, which is usually interpreted as the mean of all possible split-half coefficients.

The values estimated for this study were the test-retest reliability and the internal consistency reliability. As previously mentioned, the measurement of these types of reliability was selected for this study because they are best used for skills that are constant over time, such as intelligence. The test-retest coefficient is measured by administering a test twice at two different points in time. This type of reliability assumes that there will be no change in the quality or construct being measured. In most cases, reliability will be higher when little time has passed between tests.

The test-retest method shows the stability of an assessment instrument over time. It is also important to mention that in a test–retest reliability process there will always be some degree of error, no matter if it was applied with no sign of intervening factors because there is a strong chance that subjects will remember some of the questions from the previous test and perform better. Since perfection is impossible, most researchers accept a lower level, either 0.7, 0.8 or 0.9, depending upon the particular field of research (Traub, 1994, p.89).

Regarding internal consistency, as we mentioned before, it measures whether several items that propose to measure the same general construct produce similar scores. The common rule to interpret an internal consistency value is the following:

Table 3: Cronbach's alpha values for interpreting internal consistency.

Cronbach's alpha	Internal consistency
$\alpha \geq .9$	Excellent
$.9 > \alpha \geq .8$	Good
$.8 > \alpha \geq .7$	Acceptable
$.7 > \alpha \geq .6$	Questionable
$.6 > \alpha \geq .5$	Poor
$.5 > \alpha$	Unacceptable

Not only low reliability values show problems with the items, Very high internal reliabilities (0.95 or higher) may indicate that the items may be entirely redundant. The goal in designing a reliable instrument is that scores on similar items should be related (internally consistent), but each item should contribute with some unique information as well.

An alternative way of thinking about internal consistency is that it is the extent to which all of the items of a test measure the same latent variable. The advantage of this perspective over the notion of a high average correlation among the items of a test - the perspective underlying Cronbach's alpha - is that the average item correlation is influenced by skewness (in the distribution of item correlations) just as any other average is. Thus, while the modal item correlation is zero when the items of a test measure several unrelated latent variables, the average item correlation in those cases will be greater than zero. Therefore, whereas the ideal of measurement is for all items of a test to measure the same latent variable, alpha has been

demonstrated many times to attain quite high values even when the set of items measures several unrelated latent variables (Traub,1994).

5.2 Statistical Estimations.

Once the data was collected by administrating the test, the first time in October 2011 and the second time in November 2011, the information was transferred to the SPSS statistic program. It is important to mention that the first administration of Scale A, will be referred in this chapter as administration *a*, and the second administration of Scale A will be referred as administration *b*. It is also important to remember that the Scale A is divided into 8 subscales according to the 8 main mathematical concepts that this test pretends to assess.

Before estimating the internal consistency and the test-retest value, a total sum score of administration *a* and a total sum score of administration *b* were calculated and the respective frequency tables were created (Appendix C).

Then, the first coefficient to be estimated using the Cronbach's alpha measure was the internal consistency. This coefficient was calculated for:

- The whole test in administration *a*
- The whole test in administration *b*.
- The internal consistency value for each sub-scale in administration *a*
- The internal consistency value for a each sub-scale in administration *b*

The following tables show us the values obtained through the Cronbach's alpha formula:

Table 4: Internal Consistency for all items, Administration *a* and *b*.

	Cronbachs alpha	N of Items
Administration <i>a</i>	.881	40
Administration <i>b</i>	.875	40

Table 5 and 6: Internal Consistency for each sub-scale

Table 5: Administration a

	Cronbachsalph	N of ítems
a1	0.615	5
a2	0.393	5
a3	0.717	5
a4	0.769	5
a5	0.723	5
a6	0.573	5
a7	0.538	5
a8	0.491	5

Table 6: Administration b

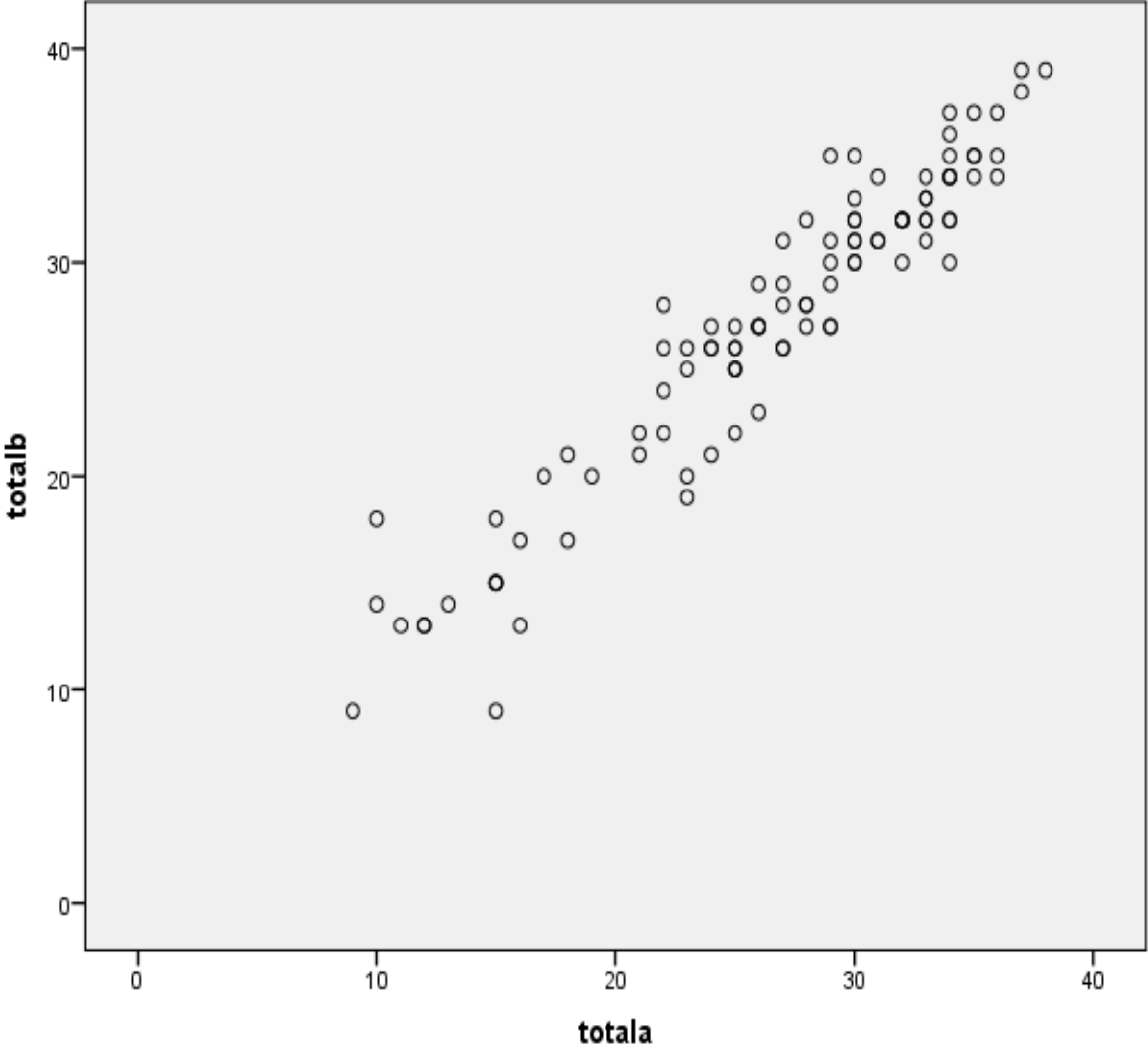
	Cronbachsalph	N of ítems
b1	0.382	5
b2	0.283	5
b3	0.584	5
b4	0.687	5
b5	0.689	5
b6	0.628	5
b7	0.692	5
b8	0.262	5

After estimating the internal consistency values, the test-retest or coefficient of stability was measured through the Pearson's Correlation.

The resulting value for the total of the scale was:

$$r_{xy} = 0.952 ; N=100; p(\text{level of significance})= 0.00$$

Figure 1: Pearson Correlation (administration *a* and *b*)



Finally, the test-retest for each subscale was measured:

Table 7: Test-retest for each subscale.

		Pearson Correlation	N	P
Sum a1	Sum b1	0.581	100	.000
Sum a2	Sum b2	0.686	100	.000
Sum a3	Sum b3	0.646	100	.000
Sum b4	Sum b4	0.747	100	.000
Sum b5	Sum b5	0.860	100	.000
Sum b6	Sum b6	0.792	100	.000
Sum b7	Sum b7	0.783	100	.000
Sum b8	Sum b8	0.707	100	.000

5.3 Discussion:

Internal Consistency

The alpha coefficient for the total of items is .881 in *administration a* and .875 in *administration b*. This suggests that the items have relatively high internal consistency. (Note that a reliability coefficient of .70 or higher is considered "acceptable" in most social science research situations.) A high value of alpha is often evidence that the items measure the underlying construct. However, if the alpha coefficient is too high, it can also suggest a high level of item redundancy; that is, a number of items asking the same question in slightly different ways.

The internal consistency for each subscale was also estimated in *administration a* and *administration b*. Comparing the tables of coefficients for the subscales (table 5 and table 6), we find that the internal consistency value for subscale number 2 is the lowest for both administrations. The low internal consistency of subscale number 2 may suggest that these groups of items are not so representative for the total of items of the Scale. The items which compose subscale number two are:

Task A6. Look at these pictures. Point out the picture with something that cannot fly.

Task A7. Look at these squares. Point out the square with five blocks but without triangles.

Task A8. Look at these pictures. Point out all the grey circles.

Task A9. Here you see people. Point out all the people who have a bag, but no glasses.

Task A10. Here you see an apple with a stalk without a leaf and a little worm coming out of the apple. [The teacher points out the apple in the square at the left top of the page.] Point out all the apples which are exactly the same.

These items are under the concept of classification in the UEMCT. According to Piaget, the acquisition of classification skills occurs through an evolving process in children age two to eleven. In this developing process we can distinguish two stages, the pre-operational stage (children from two to seven) and the concrete operational stage (children from seven to eleven). During the preoperational stage, children become able to classify objects by a single feature, for example, they become able to group together all the red blocks regardless of shape or all the square blocks regardless of color. During the concrete operational stage, children become able to classify objects according to several features and can order them in series along a single dimension such as size. In accordance to this theory, the classification skills are acquired alongside with the child's development. Therefore it is important to mention that since this test was applied to children between 4 and 5 years old who are still in a process of acquiring the classification skills, it may also be considered normal the failing in these items and in consequence a low score in this concept.

Comparing again the values in tables 5 and 6, we find that the internal consistency value in subscale number 4 is very high for both administrations, meaning that this group of items may be very representative for the whole test. The items which compose subscale number 4 are:

Task A16. Here you see squares with apples in it. Point out the square where the apples are ranked from big to small.

Task A17. Here you see squares with rocks. Point out the square where the rocks are ranked from thin to thick.

Task A18. Here you see squares with marbles. Point out the square where the marbles are ranked from small and brightly colored to big and dark colored.

Task A19. [The teacher gives the child a work paper and a pencil.] Here you see dogs. Each dog is going to get a stick. A big dog is going to get a big stick and a small dog is going to get a small stick. Can you draw lines from the dogs to the sticks they are going to get?

Task A20. Here you see slices of bread in a row from many to few slices of bread. [The teacher points out the row of slices of bread at the bottom of the page] These slices of bread fit somewhere in the row, [The teacher points out the slices of bread in the square at the left top of the page] point out where in this row these slices of bread fit in.

These items are under the concept of seriation in the UEMCT. As well as the development of classification, the development of seriation skills is also a step-by-step developmental process. As children grow and develop, their ability to seriate will also develop. For instance, a very young child (two years old) may not be able to seriate at all. The concepts of “small, smaller, smallest” and “large, larger, largest” probably will not mean much to such a young child. However, an older child (four years old) may be capable of seriating three or four objects quite easily. Just like any other skill, the ability to seriate will develop with each individual child at his/her own rate. As we can see, according to Piaget's theory, the items of subscale 4 should be perfectly suitable for the children that compose the sample of this research (they were between 4 and 5 years old), and we can see that the results confirm this idea, since for both administrations the internal consistency values are very high.

Previous researches have also shown good values of internal consistency for the Utrecht Early Mathematical Competence. A research about number sense in young children, conducted at the University of Helsinki by Pirjo Aunio in 2006, described that the Cronbach's alphas of the test indicated good and acceptable reliabilities on all the scales in the whole sample, and for children under six-and-a-half years old. The Cronbach's alpha value for the whole test in the sample was 0.79. It is important to mention that in this research, a Finnish version of the Utrecht Early Mathematical Competence Test was used.

In 2009 a study about the cognitive processes that underlie the individual differences in early mathematical performance in elementary school children was carried out by Jose Navarro, Mario Aguilar and Gonzalo Ruiz at the University of Cadiz in Spain. For this research, the Spanish version of the test was used and the value obtained for the Cronbach's alpha was also very high (0.863).

These examples suggest that good adaptations of the Utrecht Early Mathematical Competence Test, constitute solid assessment instrument, able to produce reliable results in different cultural contexts.

Test Re-Test or Stability Consistency

The stability value for the total of the scale was 0.952. A good test-retest reliability coefficient goes between a range of 0.7-1. So the Spanish version of the Utrecht Early Mathematical Competence Test shows a very high stability coefficient and this suggests that it has a good level of consistency across time. Test-retest reliability is desirable in measures of constructs that are not expected to change over time, such as intelligence. So, since the purpose of the UEMCT is to assess pre mathematical cognitive skills, a high degree of repeatability is desired for this test.

When it comes to interpret the stability coefficient of a test, it is not only important to consider what type of construct the test is measuring but also the length of time between the test and the re-test. (Di Vesta & Thompson, 1970). If the time interval between measurements is too short, respondents can remember their earlier responses and will appear more consistent than they actually are. Memory effects can lead to inflate reliability estimates. However, if the time interval between measurements is too long, interpretation problems may also appear.

After a long period between the test, the student may have acquired new knowledge that will enhance his performance on the second test and therefore he or she will obtain higher scores.

5.4 Conclusion

The aim of this study was to measure the reliability of the Spanish version of the Utrecht Early Mathematical Competence Test (Scale A) in an Ecuadorian kindergarten environment. This version was developed for Spanish children and therefore it is important to determine how reliable the results of this test are in a Latin American context. Currently, Ecuador has no assessment tool that is standardized for measuring young children's number sense, and mathematical interventions have also been lacking. Thus, it is very important to count with a reliable and valid instrument that helps pedagogues, psychologists and other professionals in Ecuador with the process of preventing future mathematical learning problems in schools.

According to the results obtained from this research, the Spanish version of the UTMC (Scale A) shows a high level of stability and a high level of internal consistency. This leads us to infer that the results of the Spanish version are reliable within an Ecuadorian environment. However, it is important to remember that in order to generalize these conclusions, future studies based on the same type of sample and following the same methods should be done. It is also important to mention that reliability is only one of the criteria that determine the quality of a test. So this study gives rise to other possible researches where quality indicators such as the validity of the Spanish Version of the UEMCT can be measured inside a Latin America context.

Another study that would be necessary to determine the adaptability of the UEMCT in Ecuador or Latin America, is a cross cultural study that analyses how the differences between the use of Spanish language in Latin American and the use of Spanish language in Spain influences the understanding of the test items. Also a replication study of this research would be worthwhile, in order to confirm the results or to inspire new research combing previous findings from related studies.

The major limitations in the study were the potential factors or threats that were able to weaken the validity of the obtained data. Therefore it was important to take into account the

objectivity of the UEMCT test during the administrations of it. The objectivity of a test refers to whether its scores are undistorted by biases of the individuals who administer and score it. Some tests, especially essay tests have low objectivity because the conditions of administration and scoring are very flexible. Tester bias can occur easily under these conditions. In contrast, multiple-choice or other closed –form tests generally are much more objective, because they are mostly self administered and all scores can apply a scoring key, which allows them to agree perfectly. A well developed test will include a manual that indicates the steps that should be followed for any situation that might affect an individuals' test performance. It is important the developers of the test specify the time that the administration of the test should take, if the instructions can be repeated, how much personal interaction should exist between the tester and the test taker and how to work with the scoring procedures, including situations such as when the test taker marks two choices on a multiple-choice task. In the case of the Utrecht Early Mathematical Competence Test, we can say that it is an objective test: all of the tasks have just one correct answer; most of the items are multiple-choice form and the instructions of administration and of the score process are well explained in the test manual written by the developers. There were some uncertainties about how to score the test in case of special circumstances, but communication by e-mail with, Hans Van Luit, one of the developers of the UEMCT was established and he clarified all the doubts. So the objectivity of the test itself was not a threat to the validity of this study. It is the aim of measuring the reliability of the test scores what represented itself a threat for the study. The reliability of a measurement system is defined as the ability of the system to perform its required function, within specified working conditions for a stated period of time. Unfortunately, factors such as manufacturing tolerances in an instrument and varying operating conditions conspire to make the faultless operating process of a system impossible to predict. Such factors are subject to random variance and chance, and therefore reliability cannot be defined in absolute terms. However, for this study, conditions such as the time between both administrations of the test, the place where it was administered, the time of the day when it was administered and other factors were planned in a very cautious and detailed way in order to avoid as much as possible any external influence in the performance of the test takers.

Before concluding this dissertation, some recommendations will be presented bellow regarding the items of the UEMCT. According to the results obtained by estimating the internal consistency for each subscale, we can recommend:

- A revision considering possible adjustments or changes in the 5 items of subscale number 2. The internal consistency for this group of items was the lowest value in comparison with the coefficient for the other subscales and this may suggest the items on the test measure diverse knowledge or skills than those which the test is supposed to measure.
- A very high value of cronbachs alpha can possibly indicates that the there is a redundancy of items and therefore some items should be deleted. The value of internal consistency for the whole test in this research was .881 in *administration a* and .875 in *administration b*. In the field of social research, these are acceptable values, however if due to reasons of efficiency, professionals might want to reduce the number of items of the Spanish version of the UEMCT, we recommend the construction of a shorter version of the test taking as a base the items of subscale number 7 which turned to be the most representative group of items according to the results of the internal consistency estimation.

It is important to emphasize that the recommendations presented above are directed to the psychologists and researchers who are qualified to work with the constructions of tests into the psychometric field and that they were suggested based only in the results of the present research. It is also important to mention that they do not represent the aims of this research but may constitute the focus of future studies.

As we can see the present study is intended to be a contribution into the research field of the mathematical learning process and into the field of early intervention. There is an increasing worldwide interest in understanding how mathematical comprehension occurs and this is leading to investigations aimed to develop the knowledge about prevention and intervention of mathematics difficulties.

Literature reference list

- Ansari, D., Donlan, C., Thomas, M. S. C., Ewing, S. A., Peen, T., & Karmiloff-Smith, A. (2003). What makes counting count? Verbal and visuo-spatial contributions to typical and atypical number development. *Journal of Experimental Child Psychology*.
- Barouillet, P. & Camos, V. (2006). Mathematical cognition in children. Solal.
- Berch, D. & Mazocco, M. (2007). Why is it Math so hard for some Children?: The nature and origins of Mathematical Learning Difficulties and Disabilities.
- Bideaud, J. & Lehalle, H. (2002). The development of numerical activities in children. Lavoisier.
- Cheeseman, J., & Clarke, B. A. (2007). Young children's accounts of their mathematical thinking. Adelaide SA: MERGA Inc.
- Dehaene, S. (2001). The cognitive neuroscience of numeracy: Exploring the cerebral substrate, the development, and the pathologies of number sense. In S. M. Fitzpatrick & J. T. Bruer (Eds.), *Carving our destiny: Scientific research faces a new millennium* (pp. 41-76). Washington, DC: Joseph Henry Press.
- Desoete, A. (2007), Students with Mathematical Disabilities in Belgium: From Definition, Classification and Assessment to STICORDI Devices, in Thomas E. Scruggs, Margo A. Mastropieri (ed.) *International Perspectives (Advances in Learning and Behavioral Disabilities, Volume 20)*, Emerald Group.
- Di Vesta, F. & Thompson G. (1970). *Educational psychology*. New York, USA. Appleton-Century-Crofts.
- Dockrell, J. (1998), *Development of mathematical skills*. England.
- Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Holland. Reidel Publishing Company.
- Gall, M.D., Gall, J.P. & Borg, W.R. (2007) *Educational Research. An Introduction*. New York: Longman Publishers. 8th Ed.
- Garcia- Madruga, J.A. (1992). Introducción al procesamiento conectivista. Introduction to the connectivist process. "*Math Learning Disabilities*." *Division for Learning Disabilities Journal of CEC*. Madrid: Alianza Editorial Garnett, K.

Garhart, C.(2000). Theories of Childhood: An Introduction to Dewey, Montessori, Erikson, Piaget & Vygotsky. USA. Redleaf press.

Garnett K. (1998). Number sense in children. *Division for Learning Disabilities Journal of CEC*. USA.

Gellistal, R. C., & Gelman, R.; (1978).The child's understanding of number.USA, President and Fellows of Harvard College.

Geary, D., Hoard, M.K, and Hamson, C.O. (1999). Numerical and arithmetical cognition: patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology*.

Graham, A., Lyn, D., &Bussi, M., (2008). Handbook of International Research in Mathematics.USA, Taylor and Francis.

Griffin, S., Case, R., &Siegler, R. (1994).Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at-risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice*. Cambridge, MA: Bradford Books, MIT Pres

Groth-Marnat ,G. (2009). Handbook of Psychological Assessment, United States of America, John Wiley & Sons.

Kline, Paul (1986) A Handbook of Test Construction, introduction to psychometric design, Great Britain, University Press, Cambridge.

Koontz, K. L., &Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children.

Kosc, L. (1974) Developmental dyscalculia. *Journal of Learning Disabilities*

McCloskey, M. (1992), "Cognitive Mechanisms in Numerical Processing: Evidence from Acquired Dyscalculia,".

NJCLD (2010).National Joint Committee on Learning Disabilities. Retrieved from:
<http://www.ldonline.org/about/partners/njclld>

Popham, James (2003), Test better, Teach better: The instructional role of assessment. USA, Association for Supervision and Curriculum Development.

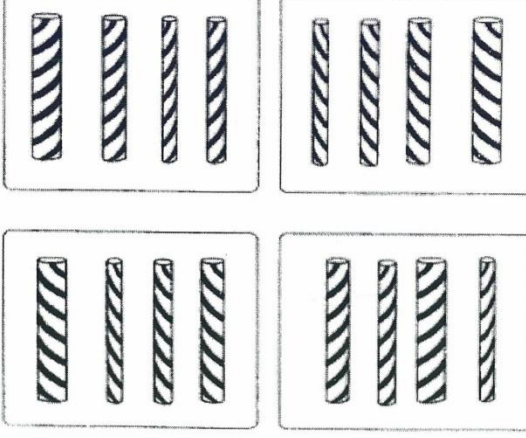
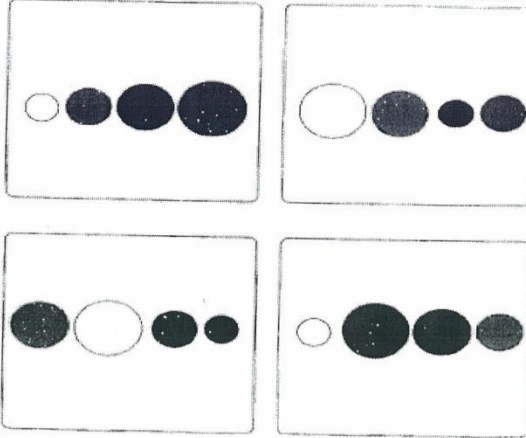
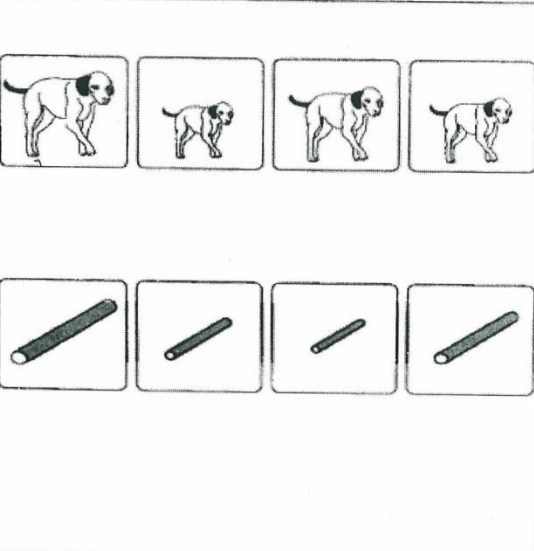
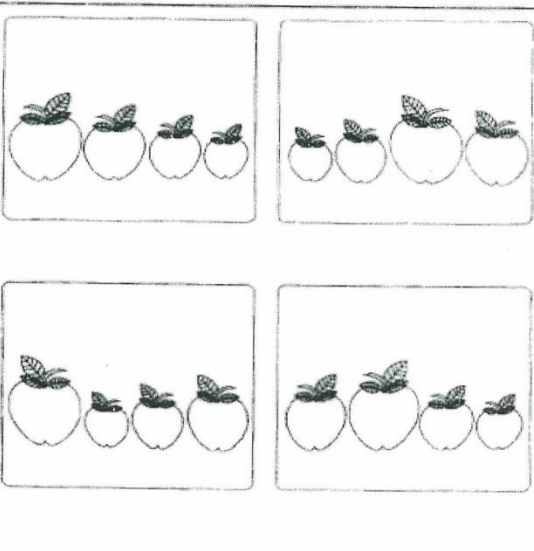
- Ravid, Ruth (2011), *Practical Statistics for Educators*. United Kingdom, Rowman and Littlefield, Publishers.Inc.
- Romberg, T., (1992).*Mathematics Assessment and Evaluation*. USA. States University of New York.
- Rumelhart, D. E., McClelland, J. L., and the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1. Cambridge, MA: MIT Press.
- Rust, J. (2009). *Modern psychometrics: the science of psychological assessment*. London, Great Britain. International Ltd Padstow.
- Saccuzzo D., Klapan R. M, (2009) *Psychological Testing: Principles, Applications, and Issues*. 7th Edition.
- Schweingruber, H., Woods, T., & Cross, C., (2009). *Mathematics learning in early childhood*. USA, National Academic Press.
- Sharma, M. C. (1986), *Dyscalculia and other Learning Problems in Arithmetic: a historical perspective*. Focus on Learning Problems in Mathematics.
- Sirkin, M. (2006).*Statistics for Social Science*.USA, Sage Publications.
- The American Psychiatric Association.*The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)*. USA.
- Traub, R. (1994). *Reliability for the Social Sciences: Theory and applications*. USA, Sage Publications.
- Van Parreren, C.F. (1989). *Principles of action psychology (II)*. *Acts - Journal of Theoretical Research*. Holland.
- Wong, B., Graham L., Hoskyn, M., & Berman, J. (2008).*The ABCs of Learning Disabilities*. USA, Elsevier Academic Press, Second Edition.
- Wynn, K. (1992). *Children's acquisition of the number words and the counting system*.*Cognitive Psychology*.

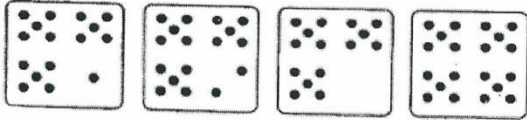
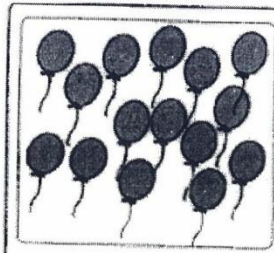
Yael Onn, (2005) Privacy in the Digital Environment, Israel, Haifa Center of Law & Technology.

Zeller, R., Carmines, E. (1980). Measurement in the social sciences: the link between theory and data. USA, Cambridge University Press.

Appendix

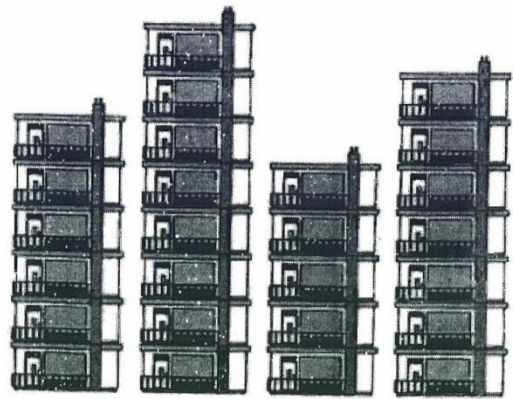
Appendix A: Sample of the UEMCT cards.

	
<p>a b c d</p> <p style="text-align: right;">A 17</p>	<p>a b c d</p> <p style="text-align: right;">A 18</p>
	
<p>voor kopieerblad, zie achterin.</p> <p style="text-align: right;">A 19</p>	<p>a b c d</p> <p style="text-align: right;">A 16</p>



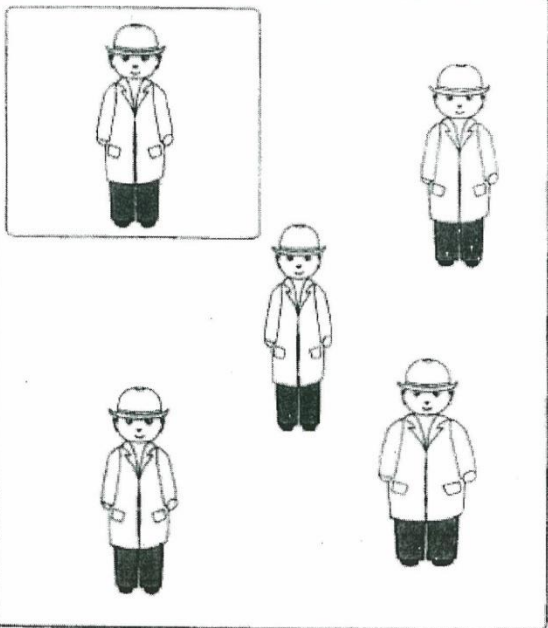
a b c d

A 15



a b c d

A 3



c b a d

A 2



a b
c d

A 6

Appendix B:**Scale scoring table: from test score to competence score**

Test Score	Version A	Version B
0	0	0
1	0	0
2	10	16
3	16	21
4	21	26
5	25	29
6	28	32
7	31	34
8	34	37
9	36	39
10	38	41
11	40	42
12	42	44
13	44	45
14	46	47
15	48	48
16	49	50
17	51	51
18	52	52
19	54	53
20	55	55
21	57	56
22	58	57
23	59	58
24	61	60
25	62	61
26	64	62
27	65	64
28	67	65
29	68	66
30	70	68
31	71	69
32	73	71
33	75	73
34	77	75
35	80	77
36	83	80
37	86	84
38	91	89
39	100	98
40	100	100

Appendix C:

Figure 2: Frequency distribution of the raw scores in administration *a*.

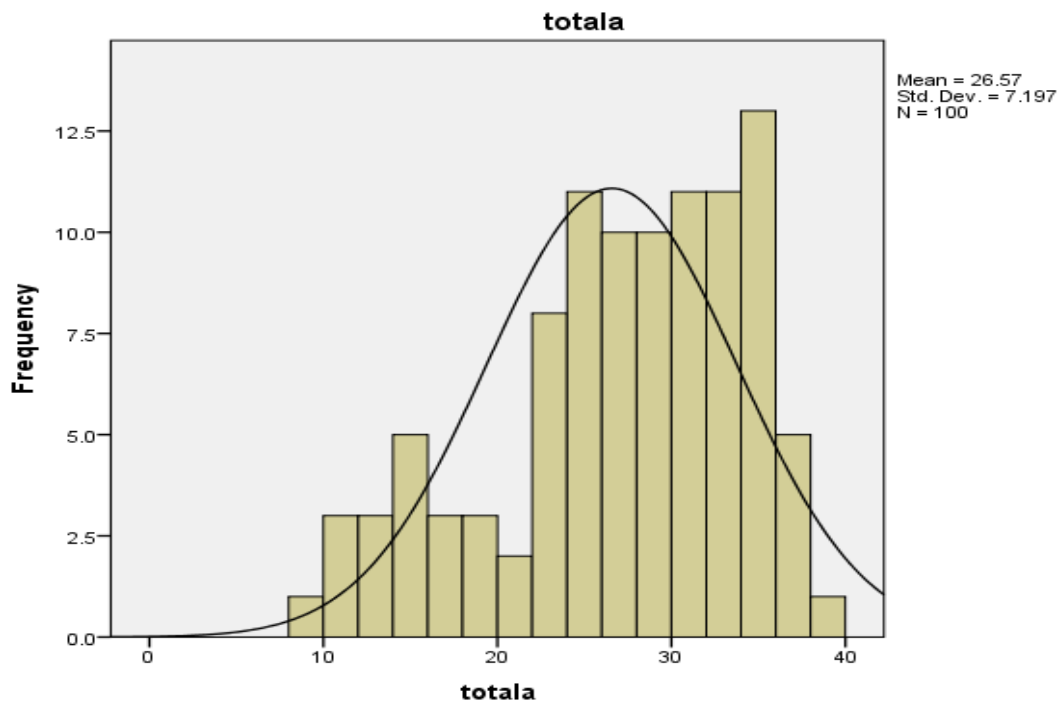


Figure 3: Frequency distribution of the raw scores in administration *b*.

