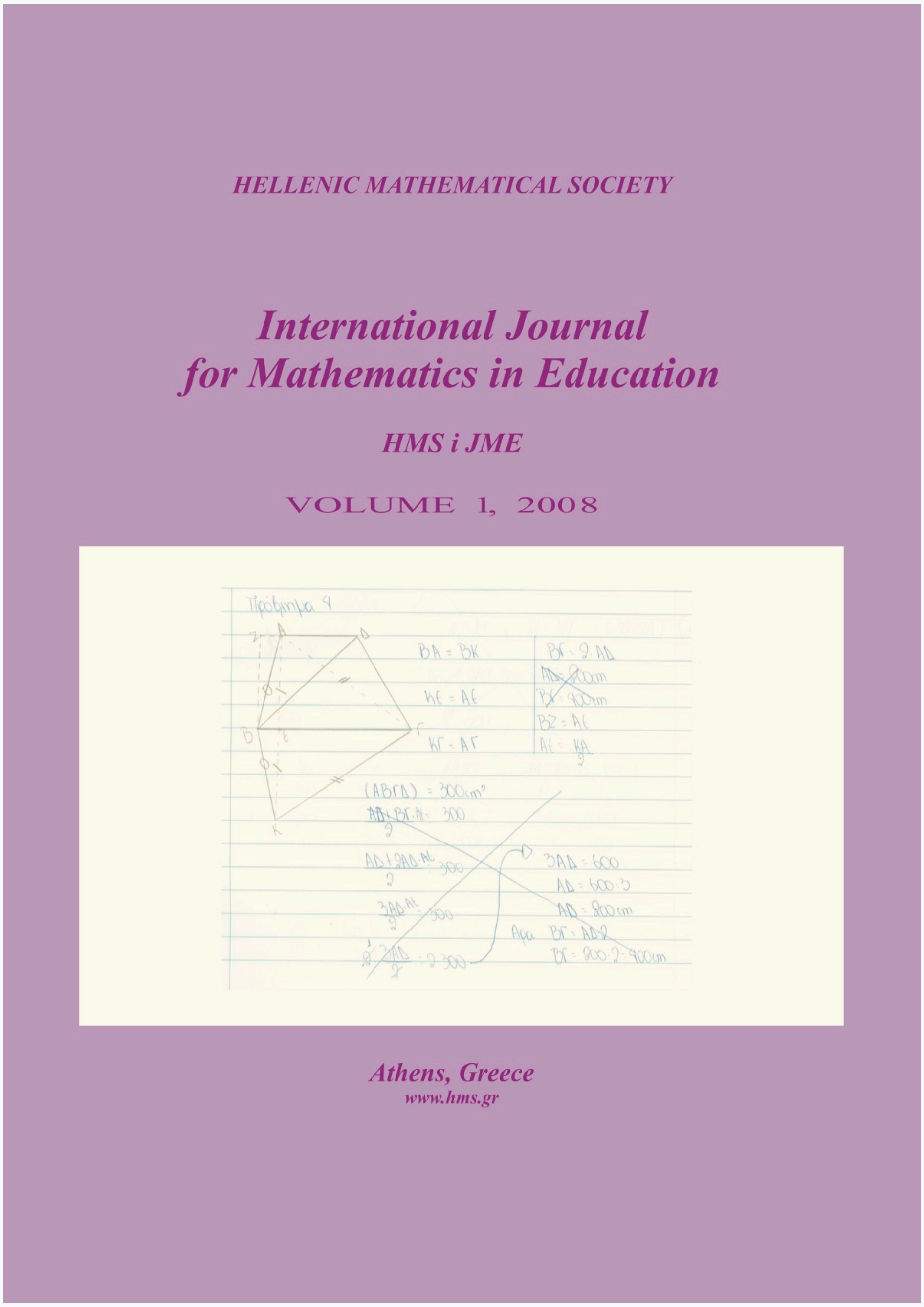
International Journal for Mathematics in Education

Vol 1, No 1 (2008)

International Journal for Mathematics in Education HMS i-JME

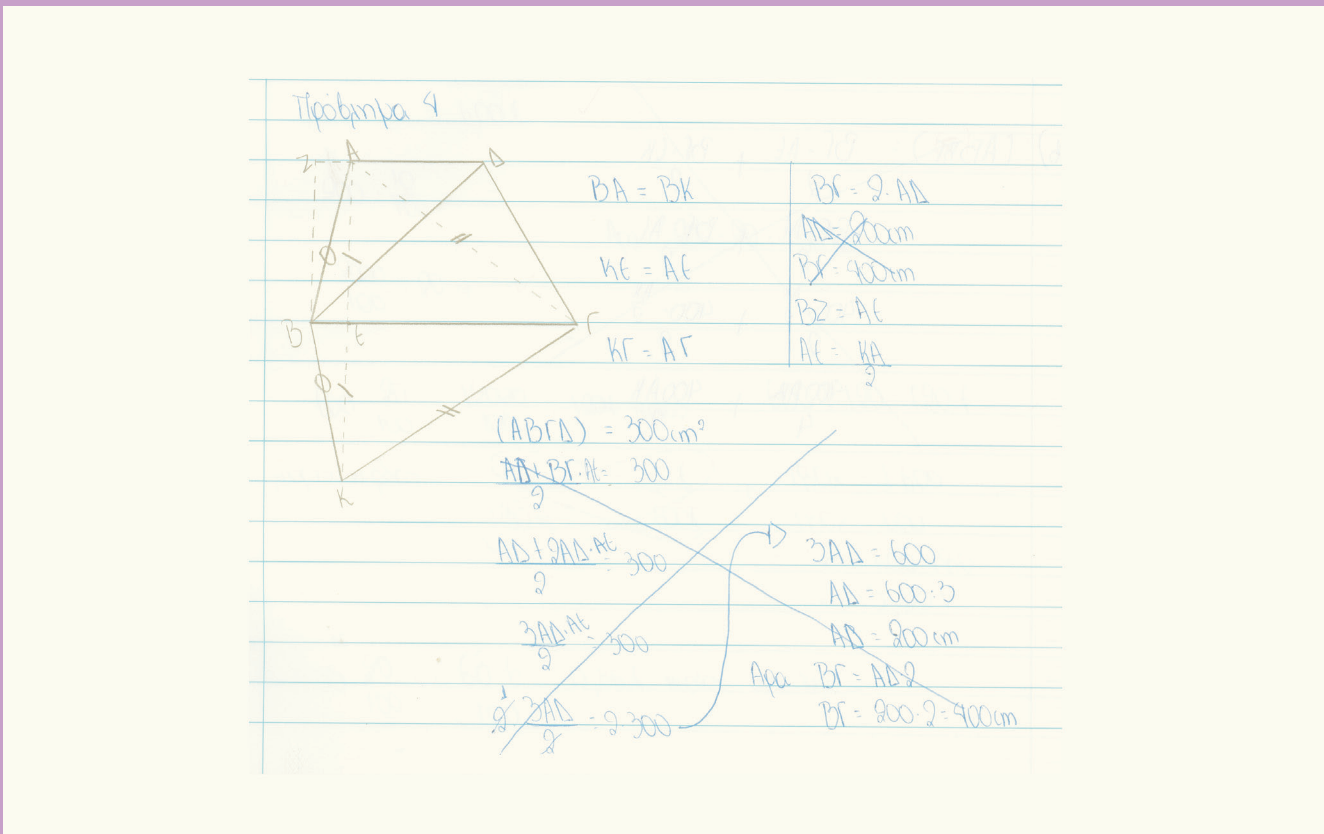


HELLENIC MATHEMATICAL SOCIETY

International Journal for Mathematics in Education

HMS i JME

VOLUME 1, 2008



Athens, Greece
www.hms.gr

Contents

ISSN: 0072 - 7466

GREEK MATHEMATICAL SOCIETY
34, Panepistimiou (Eleftheriou Venizelou) Street, GR. 106 79 - Athens - HELLAS
Tel. +30 210 3616532 - +30 210 3617784 - Fax: +30 210 3641025
e-mail : info@hms.gr - www.hms.gr

International Journal for Mathematics in Education (*HMS i JME*)

The Hellenic Mathematical Society (HMS) decided to add this Journal, the seventh one, in the quite long list of its publications, covering all aspects of the mathematical experience. The primary mission of the HMS International Journal for Mathematics Education (***HMS i JME***) is to provide a forum for communicating novel ideas and research results in all areas of Mathematics Education with reference to all educational levels.
The proposals must be written almost exclusively in English but may be admitted, if necessary, also in French, in German or perhaps in Spanish.
The proposals could concern: research in didactics of mathematics, reports of new developments in mathematics curricula, integration of new technologies into mathematics education, network environments and focused learning populations, description of innovative experimental teaching approaches illustrating new ideas of general interest, trends in teachers' education, design of mathematical activities and educational materials, research results and new approaches for the learning of mathematics.

Subscription The annual subscription rate is 15 euro.

Address: *Hellenic Mathematical Society*
34, Panepistimiou (Eleftheriou Venizelou) Street
GR. 106 79 - Athens Greece e-mail : info@hms.gr
Tel. (0030210) 3616532 - (0030210) 3617784 - Fax: (0030210) 3641025

Hellenic Mathematical Society
International Journal for Mathematics in Education

NOTES FOR AUTHORS

Authors are asked to submit their manuscript in electronic form via e-mail to: **info@hms.gr; kalabas@rhodes.aegean.gr; kafoussi@aegean.gr,** with the indication **for hms i jme.**

The first page should contain the author's e-mail address and keywords. Papers should be written in English but also, if necessary, in French, in German or in Spanish.

The format of the manuscript: Manuscripts must be written on A4 white paper, double spaced, with wide margins (3 cm), max 20 pages, Times New Roman 12pt. Each paper should be accompanied by an abstract of 100 to 150 words. References cited within text (author's name, year of publication) should be listed in alphabetical order. References should follow the APA style (http//:www.apastyle.org/pubmanual.html).

The submitted papers are reviewed by two members of the International Scientific Committee (blind review). Two positive reviews are necessary for a paper to be accepted.

Authorship

Submission of a manuscript implies that the work described has not been published before; that it is not under consideration for publication elsewhere; that its publication has been approved by all coauthors. If the manuscript is accepted for publication the authors agree to automatically transfer the copyright to the publisher.

Hellenic Mathematical Society
34, Panepistimiou (Eleftheriou Venizelou) Street
GR. 106 79 - Athens - Greece
Tel. 3616532 - 3617784 - Fax: (0030210) 3641025
e-mail : info@hms.gr

Introduction

The first volume of the *HMS International Journal for Mathematics in Education* includes five research papers.

The purpose of the first paper by Iman Osta *"Teaching Mathematics for Understanding or for the Test? School Exams Molded by National Exams"* is to investigate the relationships between the mathematics tests developed by teachers in schools and the mathematics national examination tests, in a period of a major change of mathematics curricula, in Lebanon.

The second paper by Evangellos Papakonstantinou, Dimitris Karageorgos, Vassilios Gialamas *"Study on the Performance of Secondary School Students on Statistics and Analysis of the Respective Difficulties they Face"* considers the area of Applied Mathematics and tries to identify the variables affecting students' performance on the subject of Statistics, which is taught in the 3rd year of the Greek Gymnasium (14-15 year old students; last year of the junior high school).

The third paper written by Charalambos Lemonidis *"Longitudinal study on mental calculation development in the first two grades of primary school"* presents the results of a longitudinal experimental teaching concerning mental calculation in the four arithmetic operations for first and second graders of Primary School.

In their paper *"Prospective teachers' more a-more b solutions to area–perimeter, median- bisector tasks",* Pessia Tsamir & Demetra Pitta-Pantazi discuss the impact of the intuitive rules on Cypriot prospective elementary school teachers' solutions to perimeter and area comparison tasks.

Finally, Corneille Kazadi by his article *«Pratiques de l'évaluation formative des apprentissages: une étude de cas multiples menée auprès d'enseignants des mathématiques du secondaire»* is trying to identify the difficulties of the teacher's evaluation practice on the secondary school specific situations.

Hellenic Mathematical Society
International Journal for Mathematics in Education

Teaching Mathematics for Understanding or for the Test? School Exams Molded by National Exams

Iman Osta

Abstract

The purpose of this study is to investigate the relationships between mathematics tests developed by teachers in Lebanese schools, the math national examination tests, and the math curriculum standards, in a period of major reform.

Different categories of tests were analyzed and compared: a sample of official exam tests under the old curriculum, a sample of model tests under the new curriculum, and the math school exam tests for grades 7 and 9, in a representative sample of schools in Beirut and its suburbs. The tests were analyzed according to their mathematical content and the mathematical abilities they address. For the latter, the analysis adopted the three mathematical abilities of the NAEP framework: procedural knowledge, conceptual understanding and problem solving. The Pearson Product-Moment coefficient was used for comparisons.

Findings showed that the grade 9 school exams are still more aligned with the old official exams than with the new curriculum, four years after its implementation. The new curriculum in this case could not play the role of catalyst of change in teachers' teaching and assessing practices.

Introduction

The purpose of this paper is to investigate the relationships between the mathematics tests developed by teachers in schools and the mathematics national examination tests, in a period of major reform of mathematics curricula. The case of Lebanon is considered. The main research questions are: Does a reform of curricula automatically modify the teaching and assessment practices of teachers and make them coherent with the new curriculum? Do schools and teachers teach mathematics for understanding, as stated and advocated by the new curricula, or for preparing students to the national examination tests? What if the national examination tests do not reflect the new philosophical, pedagogical and didactical guidelines of the reformed curriculum?

More specifically, the paper attempts to show that national examinations that were in effect for a long period of time, which have adopted a stable format and set a specific assessment "culture", can continue to mold mathematics teaching and school assessment. They can continue to orient them towards preparing students to the test rather than towards achieving the goals announced in the curricula that emphasize understanding, critical thinking and problem solving. In such a case, national examinations can be considered to be an impediment to educational and curricular change, if they don't align with the new goals and approaches of the reformed curricula.

Assessment is a central operation within the educational process. For many decades, the main and probably the sole tool for assessment was formal examination, using tests. Assessing has long been confused with testing or evaluation. Nowadays, trends around the world indicate a shift toward more global forms of assessment. The term *assessment* is defined by NCTM (National Council of Teachers of Mathematics NCTM, 1995), as "the process of gathering evidence about a student's knowledge of, ability to use, and disposition toward mathematics and of making inferences from that evidence for a variety of purposes". That evidence may take many forms, among which are investigative projects, performance tasks, portfolio, problem solving situations, etc. According to the National Center for Research in Mathematical Sciences Education (1993), "Testing can be contrasted with assessment in that it involves creating a situation that will inform decisions".

Despite the fact that nowadays more global and varied forms of assessment are widely advocated, testing is still a dominant tool for assessing students' work in a majority of educational systems. Beside their function of assessing students' work and determining whether they have achieved the expectations from their learning, many research studies have shown that the examinations affect widely the teaching practices. Boud (2000), for example, contends that assessment achieves the "double duty" of judging achievement and transmitting what we value. In the same sense, O'Day and Smith (1993) advocate that "assessment is important because it is believed that what gets assessed is what gets taught".

Researchers at the American National Center for research on Mathematical Sciences Education (NCRMSE) examined the assessment procedures used by American teachers and their test-preparation practices. They studied the alignment of teacher-prepared tests with the NCTM (1989) Standards. Their results demonstrated that the assessment procedures commonly used in schools were not only inadequate but should be viewed as a major barrier to the reform of school mathematics (Romberg, 1992).

Context and Background of the Study

In Lebanon, even though various assessment tools are occasionally used in classrooms, schools adopt students' grades on tests (quizzes, short tests, midterm and final exam tests) as the main tool for decision making on students' promotion from one grade level to another, or for graduation from school. Final school exams have usually the greatest weight in such decision making. However, the major means for quality control and for certification in Lebanon is the national examination, referred to as official exams, based on nation-wide tests in various disciplines set by committees under the

Ministry of Education and a governmental body (CERD: Center for Educational Research and Development).

A national curriculum is in effect in Lebanon, which is binding to both, public and private schools. The two educational sectors, public and private, are almost equally spread in Lebanon, each one of them catering for almost half the population. Public schools are run by the Ministry of Education, while private schools are either run by individuals, groups, foundations, or religious associations. While public schools implement only the national curriculum and textbooks, the private schools are free to implement more than one program, and may use different series of textbooks, but are bound to teach the national curriculum and prepare their students for the official examinations. Official exams take place in Lebanon every year at two grade levels: the end of the intermediate cycle of study (grade 9), for the “Brevet” certificate, which gives access to secondary school, and the end of secondary level (grade 12), for the “Baccalaureate” certificate and graduation from pre-university education.

The official exams in Lebanon are high-stakes exams and have an imposing power. The results of those tests are frequently and heavily used, not only to evaluate individual students' achievement, but also to evaluate their teachers and their schools. Private schools gain and measure their reputation by the percentage of success of their students in the official exams. These percentages are used in advertising for the schools and recruiting students. Raising their students' test scores becomes then a major goal for schools, as well as an indicator of school improvement. This leads to the observed fact that teachers teach to the test, and that school administrators shape their school policies and focus their academic activities around that goal. As a result, the official exams set the standards and bring, each time, a message to all agents involved in the educational process, about the priorities of the subjects studied, about what is considered to be essential knowledge under the discipline and what is less important. They determine what is the valued mathematics that should be taught and how it should be taught.

A reform process of the curricula started a few years ago in Lebanon, after a stagnation of around thirty years during which an older curriculum was used, with no possibility to reform it, partially because of the many years of war that hit the country. The new math curriculum adopts goals and learning objectives targeting higher-order levels of thinking rather than the lower-order objectives targeting memorization of facts and learning of algorithms emphasized in the old curriculum. Constructive teaching methods and active learning of students are encouraged, as opposed to the traditional lecture-type teaching under the old curriculum. Evaluation guidelines in the new curriculum focus on assessing a more global body of abilities, including knowledge of facts, understanding of concepts, problem solving, critical thinking, rather than solely knowledge of facts and application of procedures, as under the old curriculum.

The new curriculum was gradually implemented starting 1998. The year 2000 witnessed both, its full implementation at all grade levels and the first official exams under the new reformed curriculum. Today, many years later, there is a general feeling that the new official exams have not changed enough to reflect the drastic changes in the curriculum and that the educational "culture" nurtured by the long-lived old curriculum and its official exams is still influencing the new official exams. Knowing that the power of the official exams is even more compelling to teachers and schools than the texts of the curriculum and its guidelines, it is legitimate to assume that they are setting the teaching and testing practices in schools back to the more traditional ones.

The assumption is that the extremely procedural and directive nature that has always characterized the mathematics official exams, has established a deeply rooted teaching and assessment "culture" that is not easy to modify by the mere implementation of new curricula. What are the elements of that "testing culture" developed under the 30-year lasting official exams in

Lebanon and how are they still affecting the new exams? How do teachers test their students' learning under the new curriculum? Do their school tests reflect the drastic change in the curricula? Do they reflect it equally at all grade levels, and more specifically at the beginning and the end of a cycle? Or do they still follow the same schemes set by the long-lasting official exams before the reform? What are the types of mathematical abilities that they emphasize and consequently value? Do they reflect the shift from a focus on procedures and algorithms to a focus on understanding, reasoning and problem solving, as set by the new curricula (Ministry of Education & CERD, 1997, p. 289)? An assumption of this paper is that the school exams developed by teachers reflect their teaching emphasis and practices, based on the fact that teachers test their students for what they actually taught them. Thus the above questions apply as well to the teaching practices and their relation to the curriculum.

Purpose of the Study

As announced at the beginning of the paper, the purpose of this study is to investigate, in a period of major reform of mathematics curricula in Lebanon, the relationships: 1) between the mathematics tests developed by teachers in schools and the older mathematics national examination tests, and 2) between the mathematics tests developed by teachers in schools and the new curriculum. The aim is to show that school exam tests, and thus teaching of mathematics in schools, are still under the effect of the old assessment "culture" set by the older official exams, and that the new curricula could not play the role of a catalyst of change in teachers' teaching and assessing practices, towards the standards and guidelines of the new curriculum. Diagram 1 provides a visual organization of this aim.

The following steps are undertaken:

1. Analysis of the mathematics school examination tests for grade 7 and grade 9, in a representative sample of schools in Beirut and its suburbs. This analysis will consider the content covered by the school exams and the mathematical abilities they address. These two

grade levels were selected for the following reason: grade 7 and grade 9 are the beginning and the end years of the Intermediate cycle, by the end of which students sit for the Brevet official exams. Considering these grade levels will provide a picture of the tendencies in mathematics teaching and school assessment throughout the intermediate level.

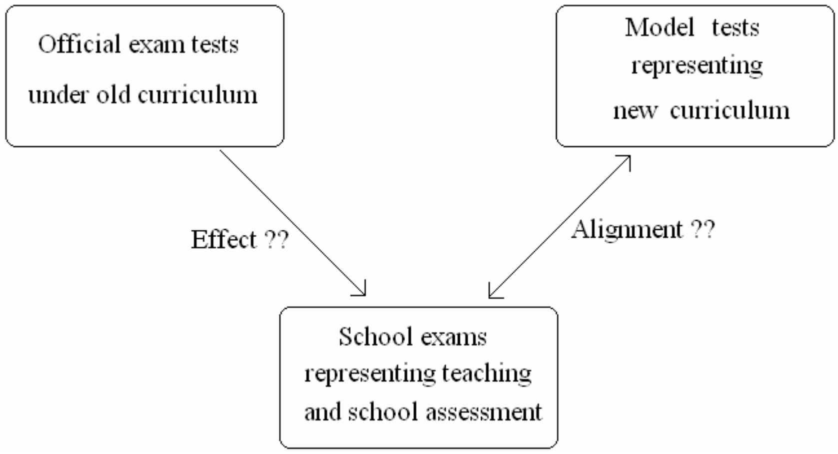


Diagram 1. *Research questions and components involved*

2. Comparison between the grade 9 school tests for the year 2004 and the official exam tests that were in effect just before the reform (throughout six years before the year 2000). The aim is to explore whether the school exams in the year 2004 are still affected by the official exams under the old curriculum.
3. Comparison between the grade 9 and grade 7 school tests and model tests provided in the Evaluation guide issued as addendum to the new curriculum (Ministry of Education & CERD, 2000). The aim is to explore whether the school tests in grade 7 are more aligned with the curriculum than the grade 9 school exams, because of the effect on the latter of the testing "culture" set by the official exams.

Method

The method adopted in this research is the text analysis method. The texts analyzed include: a) The objectives of the new mathematics curriculum, for the grades 8 and 9 (Ministry of Education & CERD, 1998 & 1999), which determines the content covered by the official exams, b) five model tests provided in the Evaluation guide issued as addendum to the new curriculum (Ministry of Education & CERD, 2000), c) a sample of the official exam tests in mathematics at the intermediate level, and d) a sample of the school final exam tests in mathematics, for grade 7 and grade 9, collected from a sample of schools in Beirut and its suburbs.

Sampling

Sampling of the official exam tests. Eleven tests are considered, administered over six years, the years just before the reform process started. Those 11 tests include 6 regular exam tests usually administered in June, at the end of the school year, and 5 "second-session" exams, administered in September and given as a second chance for students who did not succeed in the regular exam. In one of the years, only one session was offered.

Sampling of the school final exam tests in mathematics. A sample of Lebanese schools from the Beirut region was selected, using stratified sampling. The size of the sample is set to be 20 % of the school population size. Within each stratum, random computerized sampling was used. The types of schools were the basis for the various strata. Accordingly, the sample included: 10 public schools, 10 schools affiliated with local associations, 3 schools affiliated with religious missions, and 8 schools owned by individuals or groups of individuals. From each one of those schools, the year 2004 mathematics final exam tests for grades 9 and 7 were collected.

Analysis techniques

The study is quantitative, descriptive and comparative. It uses simple statistics such as percentages and correlation coefficients (namely Pearson Product-Moment coefficient).

The objectives of the mathematics curriculum for grades 8 and 9 were organized and coded, for the purpose of this study, under five subject strands: Numbers, Algebra, Measurement, Analytic geometry, and Geometry (Euclidean geometry), numbered I to V respectively. Under each subject strand, the curriculum content was organized in a hierarchy of objectives and sub-objectives that were coded accordingly. For example, the code II.5.2.2.b designates the sub-objective "Verify whether a given pair is a solution for a system of two equations in two unknowns", under the objective II.5.2.2. of the topic 5.2. "Systems of first-degree equations in two unknowns", under the subject strand II (Algebra).

In order to analyze the levels of cognitive abilities that the various tests address, we have adopted the three mathematical abilities included in the American 1990-2003 National Assessment of Educational Progress (NAEP) framework: Procedural Knowledge (P.K.), Conceptual Understanding (C.U.), and Problem Solving (P.S.). These abilities are globally defined (National Center for Education Statistics, 2005) as follows:

- *Procedural knowledge is often reflected in a student's ability to connect an algorithmic process with a given problem situation, to employ that algorithm correctly, and to communicate the results of the algorithm in the context of the problem setting.*
- *Conceptual understanding reflects a student's ability to reason in settings involving the careful application of concept definitions, relations, or representations of either.*
- *Problem-solving situations require students to connect all of their mathematical knowledge of concepts, procedures, reasoning, and communication skills to solve problems.*

For each one of the abilities, the NAEP document provides more specific and detailed specifications. These specifications were used in this study for the purpose of classifying the test items of the various types of tests.

A double-entry matrix was constructed, having as first entry (row headings) the list of objectives and sub-objectives of the curriculum coded as explained before, and as second entry (column headings) the set of three mathematical abilities (P.K., C.U. and P.S.). Four such double-entry matrices were used to map the test items according to the mathematical content and the mathematical abilities they address: one for official exams and one for model exams, and two matrices for school final exams (one for grade 7 and one for grade 9).

Definition of a test item. In objective-type tests, the test items are usually independent and isolated from each other. Their categorization is easy and clear-cut. This is not the case for the tests analyzed in this paper. They are formed of more open questions with interrelated subquestions, which makes it a must to develop criteria for classifying them, and to define what is to be considered as one test item, since the number of test items is the basis for calculating the percentages for comparison. We define a "test item" as being any part of the test that requires a response from the student, which entitles him/her to a part of the grade. A test item may take one of the two following forms:

- A question that requires an answer. For example, "What is the nature of triangle ABC?"
- An imperative sentence, such as "Calculate the coordinates of P".

In the case of many components required in one sentence, it is considered to stand for more than one test item. For example, "Plot the points A, B, C, and the straight line (D)" is counted as four items, because it stands for "Plot point A, plot point B, plot point C, plot straight line (D)".

Procedure. Each test item (in each of the sample official, model and school exams) was solved in order to identify the knowledge content and the mathematical ability it addresses. Those items were tallied in the corresponding double-entry matrix, according to the subtopic of content as coded vs. the mathematical ability it addresses.

Since the different types of tests have different numbers of test items, the data in every matrix were transformed into percentages relative to the total number of test items in each set of tests, in order to normalize the basis of comparison. These data, organized in the four matrices, are also used to calculate the different correlations, in order to measure the relationships between the school exams and the official exams on one hand, and between the school exams and the new curriculum (as represented by the model exam tests) on the other hand.

In order to increase the validity and reliability of classification, we resorted to judging: the classification was done separately by the researcher and by an assistant researcher with a Master degree in Math Education, on a pilot sample of two official exam tests, four school exam tests, and one model test. Then the two classification schemes were compared, and the differences were discussed. Based on that discussion, the criteria for classification were refined, and then the two judges made the classification of all exams based on those criteria. The slight differences found after this second round of classification were discussed until an agreement about each one of them was reached.

Results

In order to facilitate the presentation of results, we will present the data from each one of the four matrices under the five global subject strands.

Table 1 presents the distribution of the percentages of test items in the official exams, by subject strand and mathematical ability.

As Table 1 shows, procedural knowledge was dominant in the mathematics official exams. It comes first of all three mathematical abilities, with the highest percentage (60.85 %). This means that the greatest part, around two thirds, of the official exam tests required rote application of algorithms or known procedures, privileging skills acquired by drill and practice rather than conceptual understanding. As to the share of conceptual understanding, it was only 11.67 %. The problem solving share came second, occupying almost the quarter of the test items (27.47 %). the problem-solving test items in the official exams were concentrated under geometry (26.30 % out of 27.47 %), which left only 1.17 % of problem solving items for all the other subjects.

Table 1. *Distribution of percentages of test items in official exams by mathematical ability and subject strand*

	Official Exams			
Subject	P.K.	C.U.	P.S.	Total
Numbers	4.49	3.55	0	8.04
Algebra	21.40	3.42	0.85	25.67
Measurement	0.80	0	0.32	1.12
Analytic geometry	24.84	2.24	0	27.08
Geometry	9.32	3.46	26.30	38.09
Total	60.85	11.67	27.47	100

The results show that the official exams under the old curriculum seemed to partition mathematical subjects into specialized areas for each mathematical ability. While geometry is the privileged area for problem solving, algebra and analytic geometry are the privileged subject strands for procedural knowledge. They got around 45 % out of the total of 60.86 % of procedural knowledge items (21.40 % for algebra and 24.84 % for analytic geometry).

Table 2 presents the distribution, under the new curriculum, of the percentages of test items in the model tests by subject strand and mathematical ability.

Table 2. *Distribution of percentages of test items in model tests by mathematical ability and subject strand*

	Model tests under new curriculum			
Subject	P.K.	C.U.	P.S.	Total
Numbers	7.72	12.4	0	20.12
Algebra	13.17	14.92	0.93	29.02
Measurement	1.27	0	1.23	2.5
Analytic geometry	14.03	1.47	0.57	16.07
Geometry	4.24	9.8	18.25	32.29
Total	40.43	38.59	20.98	100

We can infer from the above data that the new curriculum as reflected by the model tests emphasizes conceptual understanding and attributes to it more value. Moreover, there is a noticeable balance between procedural knowledge (40.43 %) and conceptual understanding (38.59 %). On the other hand, the share of geometry from the problem solving test items is the highest (18.25 % out of 20.98 %) of the whole percentage for problem solving. Similarly, procedural knowledge items are concentrated under analytic geometry (14.03 %) and algebra (13.17 %), while conceptual understanding is given more attention under algebra (14.92 %) and number theory (12.4 %). This latter note about conceptual understanding marks the most important change in the new curriculum, as compared to the old curriculum and its official exams, where conceptual understanding was almost neglected (11.67 % in all).

Table 3 presents the results of the analysis of school exams for grade 9, and their distribution by mathematical ability and subject strand.

Table 3. *Distribution of percentages of test items in grade 9 final school exams by mathematical ability and subject strand*

	P.K.	C.U.	P.S.	Total
Numbers	8.1	0.27	0.47	8.84
Algebra	29.88	0.05	2.48	32.41
Measurement	0.74	0	0.15	0.89
Analytic geometry	22.94	0.25	0.26	23.45
Geometry	10.68	0.16	23.57	34.41
Total	72.34	0.73	26.93	100

Table 3 shows that the subject strands that get the highest share are algebra (32.41 %), geometry (34.41 %) and analytic geometry (23.45 %). Table 3 shows also the particularly high concentration in items that test procedural knowledge (72.34 %), and the very low percentage for conceptual understanding, with less than one percent. As to problem solving, it corresponds to 26.93 % of the test items. Similarly to the official exam tests and the model tests, the school exam tests emphasize problem solving mainly under geometry (23.57 % out of 26.93 % in all for problem solving, while procedural knowledge is mainly addressed under algebra (29.88 %) then under analytic geometry (22.94 %).

We clearly notice that the percentage for procedural knowledge in the grade 9 school exams is even higher than in the old official exams, which means that schools kept attributing to it a remarkably high importance. As to the other abilities, while we notice that the shares for problem solving in each of the official exams and the school final exams in grade 9 are fairly close (27.47 % and 26.93 % respectively), the difference between the shares for conceptual understanding is noticeable (11.67 % in the official exams versus 0.73 % for the school exams), which indicates that schools tend, in the last year of the intermediate cycle, toward focusing much more on drill and practice, training students to apply algorithms and procedures than helping them to construct conceptual understanding. We can also infer that teachers, in general, still perceive their role as being to train students in preparation for the official exams' "culture", as set by the official exams under the old curriculum, which is not in line with the new curriculum.

The comparison between Table 2 (distribution of the model test items) and Table 3 (distribution of the school exam items) shows remarkable and significant discrepancies, especially in the mathematical abilities addressed. While the model tests, hence the new curriculum, focus on conceptual understanding (38.59 %), a negligible share is assigned to it in the school exams (0.73 %). The percentage of test items at the level of procedural knowledge in the school exams (72.34 %) is much higher than it is in the model exams (40.43 %). This shows that, even though the new curriculum emphasizes conceptual understanding and attributes less importance to procedural knowledge than the old curriculum and its official exams, the school exams are still prisoners of the older testing "culture" set by the official exams. The school exams under the new curriculum are still emphasizing greatly what the official exams used to emphasize, and they continue to capitalize on direct procedural knowledge for two reasons: 1) because the official exams emphasized it for a long time, which molded teachers' perception of their role and of the nature of mathematics, and 2) because this type of knowledge is more cost-effective in terms of reducing the time spent and increasing the potential grade points to earn, irrespective of students' understanding of the material.

As to the share of problem solving, the school exams and the official exams are noticeably close to each other (26.93 % and 27.47 % respectively), and are both almost equally different from the model exams (20.98 %). This difference of about 6 % between the school and the model exams' shares in problem solving may be explained by the fact that this mathematical ability is rather related to a specific subject strand, which is geometry. This subject strand has been given less attention in the model exams under the new curricula than in the old official exams (32.29 % vs. 38.09 % respectively, which is almost a difference of 6 %).

The above analysis leads us to maintain that schools actually teach to the test rather than teaching for understanding or for achieving the goals of the curriculum. It also shows that schools continued to perpetrate the same testing specifications rooted by the official exams under the old curriculum, despite of the fundamental changes in testing specifications brought by the new curriculum.

The Pearson Product-Moment correlation coefficients between these various distributions provide more evidence about the above discrepancies:

Considering the distribution of percentages of test items by both subject and mathematical ability, the correlation between grade 9 school exams and the new curriculum (as represented by the model exams) is 0.63. However, the correlation between grade 9 school exams and old official exams is 0.95, which is a very high correlation. These two correlation values indicate that grade 9 school exams are still more aligned, as to their content and the mathematical abilities they address, with the old official exams than with the new curricula.

This phenomenon is even more striking if we consider the correlations between the global distributions by mathematical abilities only, irrespective of the subject strands: it was found that such correlation between the grade 9 school exams and the model exams is 0.24, while the correlation between the grade 9 exams and the official exams is 0.99. These correlations show that the school exams in grade 9, four years after the full implementation of the new curricula, are almost identical to the old official exams in their addressing the same mathematical abilities, while they differ significantly in that respect from the new curriculum (as represented by the model tests). Table 4 summarizes these correlations in order to provide a more global and comparative view.

Table 4. *Correlations of distributions of grade 9 school final tests with distributions of official exam tests and curriculum, represented by model exams*

Correlation between	Official exam tests	Curriculum (Model tests)
	By subjects and math abilities	
Grade 9 school tests	0.95	0.63
	By math abilities	
	0.99	0.24

Do the facts exposed in the above analysis about grade 9 school exams apply to grade 7 school exams? Do teachers in grade 7 emphasize procedural knowledge as well and as much, and neglect conceptual understanding? Do they comply to the older official exams rather than with the new curriculum? Following are the results of analysis of grade 7 school exams. Table 5 presents the distributions by subject and mathematical ability of the test items in grade 7 school exams.

The discrepancies that Table 5 shows between the distribution of grade 7 school test items by subject, and the distribution of the official exams and the grade 9 school exams is logically understandable and can be interpreted by the difference of math curriculum content between grade 7 and grade 9. For example, it is known that number theory is an important topic in grade 7 because it relates to the math basics that are taught at this lower grade level, while its weight decreases in higher level, to give room to other subjects such as analytic geometry, which gets in grade 7 only 2.17 % of the school exams, reserved to the one-dimensional analytic geometry (geometry on an axis). Indeed, Table 5 shows that the subject Numbers gets the highest percentage of test items (42.78 %), followed by algebra (28.25 %) then by geometry (22.44 %).

Table 5. *Distribution of percentages of test items in grade 7 final school exams by mathematical ability and subject strand*

Subjects	P.K.	C.U.	P.S.	Total
Numbers	22.89	18.06	1.83	42.78
Algebra	19.68	8.07	0.5	28.25
Measurement	2.96	0.16	1.24	4.36
Analytic geometry	1.76	0.32	0.09	2.17
Geometry	4.99	8.33	9.12	22.44
Total	52.28	34.94	12.78	100

However, what is noticeable and significant in Table 5 is the high percentage of test items that address conceptual understanding (34.94 %) and the relatively low percentage of test items that measure procedural knowledge (52.28 %), as compared to its counterpart in grade 9 school exams (72.34 %) and in official exams (60.85 %). As to the problem solving share in grade 7 exams (12.78 %), it is lower than both, its counterpart in grade 9 school exams (26.93 %) and its counterpart in official exams (27.47 %). This can be explained by the fact that in all those exams, problem solving is mainly concentrated under geometry, and geometry till grade 7 is limited to the study of shapes and their properties, and does not include much of geometrical proofs or problem solving.

The data in tables 5, 1 and 2 show discrepancies between grade 7 school exams, on one hand, and official and grade 9 school exams on the other hand. However, the data in tables 5 and 3 show noticeable similarities between grade 7 school exams' distribution and the curriculum (as reflected by the model exams), particularly as concerns the distribution by mathematical ability, and especially as concerns the share of conceptual understanding (34.94 % for grade 7 exams vs. 38.59 % for the model exams).

This comparison of data can be made clearer by using the correlation coefficients. Considering the distribution of percentages of test items by both subject and mathematical ability, the correlation between grade 7 school exams and the new curriculum (as represented by the model exams) is 0.59. However, the correlation between grade 7 school exams and official exams is 0.30. These two correlations are both relatively low, as compared to the correlations between grade 9 school tests and official and model tests. A part of this difference is due to the low correlation between distributions by subject. The correlations between distributions by mathematical ability confirm this fact and provide clear evidence about the alignment of school exams at this grade level with the new curriculum, as much as the mathematical abilities addressed are concerned. These correlations are: 0.93 between grade 7 school exams and the new curriculum, versus 0.61 between grade 7 school exams and official exams. While grade 9 school exam tests are more correlated to old official exam tests than the new curriculum, the opposite is true for the grade 7 exam tests. They are significantly more correlated to the new curriculum than to the official exams, in terms of the mathematical abilities they address.

Table 6 summarizes these correlations and provides a more global and comparative view.

The results in Table 6 indicate that teaching and assessment in grade 7 are more aligned with the new curriculum than they are in grade 9. This fact reflects a tendency, throughout the intermediate cycle, from teaching and testing for understanding as advocated in the new curriculum, to teaching to the test and for preparing students for the official exams which, for a long time, required procedural knowledge and rote application of algorithms.

Table 6. *Correlations of distributions of grade 7 school final tests with distributions of official exam tests and curriculum, represented by model exams*

Correlation between	Official exam tests	Curriculum (Model tests)
	By subjects and math abilities	
Grade 7 school tests	0.30	0.59
	By math abilities	
	0.61	0.93

Conclusion

This study considered the case of Lebanon, which provided a typical example of a major reform in mathematics curriculum, after a long period of stagnation and use of a traditional curriculum that privileged rote learning, drill and practice, and procedural knowledge. The imposing national-level official examinations under that curriculum contributed to setting the testing "culture" based on rote learning, and to molding teachers' and schools' perceptions and practices, as well as their views on the nature of mathematics and its learning.

The results of this study showed that the implementation of new curricula based on understanding, critical thinking and reasoning, and accompanied with massive teacher training and information sessions, could not achieve a real change in the teachers' testing practices to make them aligned with the new approaches and assessment philosophy of the new curricula. Contrarily, the results showed that as students and teachers approach the official exams especially in grade 9, end of the middle school years, school exams tend to be more and more aligned with the old official exams, hence the old curriculum approaches, whether in the content of test items or in the mathematical abilities they address.

However, at lower grade levels, two years before the year of sitting for the official exams, the school tests are aligned with the new curriculum and abide by its specifications, especially as much as the mathematical abilities it addresses are concerned.

While this does not necessarily mean that teachers consciously follow better the guidelines of the curriculum in grade 7 then neglect them in grade 9, it does indicate that mathematics teaching in grade 7 is definitely more coherent with the new curriculum and its objectives than it is in grade 9, when it becomes more coherent with the official exams as a main guide. Is this a coincidence? Or do the official exams set implicit guidelines that orient teaching and mold it by the end of the middle school, compelling teachers and schools to follow strictly these guidelines in order to survive in the educational system? Don't the high, almost full, correlations between grade 9 school exams and old official exams tell us about the power of those high-stakes exams, and the continuity of their effect, even under a new curriculum that has set completely different standards?

Beyond the case of Lebanon, this study poses the issue of national and high-stakes examinations, and their power in either fostering an educational and instructional change, or to the contrary in prohibiting such change. The results have shown that the new curricula could not play the role of a catalyst of change in the teachers' teaching and assessing practices, towards the standards and guidelines of the new curriculum. School exams remained stamped by the old official exams because the new official exams could not reflect enough the philosophy of the new curriculum, let alone that the older official exams reigned for 30 years.

Finally, this study highlights the major role that assessment, and particularly testing using high-stakes examinations, can play to make a curricular reform happen.

References

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167)

Ministry of Education & CERD (1997). Mathematics curricula. In *General education curricula and their objectives. Decree n° 10227* (pp. 287-327). Lebanon: Ministry of Education and Center for Educational Research and Development.

Ministry of Education & CERD (1998). *Curriculum of mathematics. Details of the contents of the third year of each cycle (Decree n° 10227, Date 8 May 1997)*. Lebanon: Ministry of Education and Center for Educational Research and Development.

Ministry of Education & CERD (1999). *Curriculum of mathematics. Details of the contents of the second year of each cycle. (Decree n° 10227, Date 8 May 1997)*. Lebanon: Ministry of Education and Center for Educational Research and Development.

Ministry of Education & CERD (2000). *Evaluation guide, Mathematics. Model tests for the official exams. Intermediate cycle*. Lebanon: Ministry of Education and Center for Educational Research and Development.

National Center for Education Statistics (2005). Mathematical abilities. Retrieved August 23, 2006, from: http://nces.ed.gov/nationsreportcard/mathematics/abilities.asp

National Center for Research in Mathematical Sciences Education (1993). Assessment and school mathematics. *NCRMSE Research Review*, 2 (3), Fall 1993. National Center for Research in Mathematical Sciences Education.

National Council of Teachers of Mathematics (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, Virginia: NCTM.

National Council of Teachers of Mathematics (1995). *Assessment standards for school mathematics*. Reston, Virginia: NCTM.

O'Day, J.A. & Smith, M. (1993). Systematic school reform and educational opportunity. In S. Fuheman (Ed.), *Designing coherent educational policy: Improving the system* (pp. 250-311). San Francisco: Jossey-Bass.

Romberg, T.A. (1992). *Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators*. Albany, NY: State University of New York Press.

Iman Osta

Lebanese American University, Beirut, Lebanon
iman.osta@lau.edu.lb

Hellenic Mathematical Society
International Journal for Mathematics in Education

Study on the Performance of Secondary School Students on Statistics and Analysis of the Respective Difficulties they Face.

Evangellos Papakonstantinou, Dimitris Karageorgos, Vassilios Gialamas,

Abstract

This study belongs in the area of Applied Mathematics and the main objective of this is to identify the variables affecting students' performance on the subject of Statistics, which is taught at the 3rd grade of Greek Gymnasium (first cycle of the Secondary School). The research comes to complete the results other studies, about this matter, have found. Of course, this study, to a certain extent, relying on the conclusions of the previous relevant research studies that have presented the influences one or two variables, usually, on the students' performance on Statistics.

1. Introduction

Many researches have studied the influence that has, on the students' performance on Statistics, the teaching of this object with P/C [1, 2]. Other researchers have studied the students' performance on Statistics, relation to opinion of them (students), about the practical usefulness of the Statistics on the real world [3, 4]. The effect of the efficient knowledge of the Mathematics on the students' performance (on Statistics) has been the study's object for many statisticians [5]. The influence of the social variables on the students' performance on Statistics is often examined by the statisticians [6]. On the area of the exact sciences, the gender of the students is another variable, which have scrutinized by many researchers [7].

The present research considers the conclusions of the previous ones, and of course, it studies many other variables. Some of these variables are the Students number in the Classroom, the existence Home Library, the Teaching experience of the Teachers in Statistics, the Students' Opinion about the School Book of Statistics, the Students' Opinion about the Statistics' degree of usefulness etc, totally we have studied twenty-five in number variables. Moreover, all these variables are studied simultaneously and finally a model is produced that takes in the statistically significant variables (seven ones). These variables are listed in the order of significance relative to the degree that each Interprets the phenomenon at hand (on Statistics) [8]. Also, the study looks for, to pin point areas in Statistics, where students face particular difficulties and to identify the respective reasons and the ways of confronting these difficulties. Finally, the study analyzes the respective difficulties, which the students face on Statistics. Of course, are drawn conclusions and put forward relevant suggestions in order to improve students' performance on this subject.

The research relies on questionnaires that were handed out to the students of the sample. Likewise, corresponding questionnaires were handed out to the students' teachers. For the statistical processing of the research' data used the technique of the Regression Analysis and the more recent one of the Structural Equation Modeling. Recently, the use of the mathematical and statistical methods has wide application on the broader field of psychological, pedagogical and educational research [9, 10].

The interpretations of the study's inferences were carried out with great cautions, so that these explanations to be within the bounds of the potentiality that have the statistical techniques that were used. It is known that, in recent years the fast development of the disciplines of the information science and Statistics, sometimes, create some problems to the researchers, about the explanation of the studies' results [11, 12].

Although, sometimes, educational research studies, in some measure, have relative worth, as to their results, for all that, they maintain their value, because they bring out the general tendency of the phenomena being examined [10]. Here, we think that the study's findings cover satisfactorily the targets of this research.

2. The Reasons Necessitating this Study

The reasons necessitating this study are:

(i) The absence of a relevant study, examining the teaching of Statistics at the First cycle of the secondary school (Greek Gymnasium) from this point of view (see the paragraph above).

(ii) The relatively inadequate teaching experience, in comparison with other teaching subjects (Mathematics, History etc.), since the teaching of Statistics has only been recently introduced in Western Europe's studies (in the first cycle of the Secondary School).

(iii) The current difficulties both teachers and students face with regard to the teaching of Statistics.

(vi) The existing teaching methods usually had as a reference and a starting point, i.e., the teaching of the so-called "classic disciplines" (Mathematics, History, Language etc.) for which there is adequate teaching experience for years.

(v) The prompt coverage, to the extent that is possible, of the teaching gaps and weaknesses with regard to the teaching of Statistics.

2. Brief Presentation of the Study's Theoretical Framework and the Variables Affecting Students' Performance in Statistics.

The current study focuses on the perusal of cognitive domain and does not touch on the issues concerning affective domain and the psychomotor one. The study of cognitive domain can be achieved with the aid of tests, which were given to the students, in the form of questions in four areas, namely: ι. Knowledge, ιι. Comprehension, ιιι. Applications and ιν. High Level Questions (Analysis, Synthesis and Evaluation).

For the realization of the study's objective, a sample of 588 students was used. Tests and questionnaires answered by the students, were used as research material. Relevant questionnaires were also handed out to the students' teachers.

Factors affecting students' performance in Statistics.

For the purposes of the study, a set of 25 variables was examined which, were regarded as possible causes for the students' performance in Statistics. The selection of these variables was based upon the usual methodology, which is applied to similar studies by the A.E.P.A. (Association for the Evaluation of Educational Achievement). The explanatory variables are classified in two categories ι. The students' family environment and his personal attributes and ιι. The features of the school's environment and the attributes of the educational environment.

For the implementation of the study's theoretical framework, the following procedure was devised in the stages described below.

Sampling

During the collection of samples the stratified random sampling technique was followed which, with the population's split into strata, helps to collect reliable samples at a relatively low cost. In particular, the Athens – Piraeus Urban Complex (UC) was selected for research, out of which samples were taken, taking into consideration the current, local, social and educational differences in the population of the Athens – Piraeus UC. Afterwards, using the random sampling technique, the Peloponnese region was selected among Greece's remaining geographical regions, in order to establish, whether there is a difference, between the Athens – Piraeus UC and the provinces, owing to the students' place of residence. Thus, 588 3rd grade Secondary School students were selected (April of 2001). A pilot study was also carried out with a sample of 80 students, which was useful for the better planning of the study and execution, as well as, for the wording of the questions.

3. Presentation of the study's tests

The tests which were handed out to the students, underwent all appropriate audits, in order to be deemed reliable and were particularly checked with respect to the following indices and criteria, in accordance with the provisions of international standards set by similar studies [13, 10]:

(a) Discrimination index. The tests were appropriate in order to distinguish above average students from average and below average ones.
(b) Difficulty index. All questions had a difficulty index between 0,2 and 0,8, in accordance with international standards.
(c) Reliability index. The employment of S.H.M. (Split Half Method) and the subsequent calculation of Pearson correlation ratio showed a high credibility degree (0,93).
(d) Criterion-related validity. Pearson ratio showed a high correlation (0,91) among students who had received high (low) scores at the initial (pilot) test and at the final (true) test respectively.
(e) Content validity. The test's content validity was ensured to a great extent, as it is shown on the following Secondary School's table of specifications.
(f) Marks adjustment. The study did not reveal any significant differences with respect to the element of luck, so as to necessitate the corrective adjustment of marks.

THEMATIC SECTION	**Number of Questions**				
	Knowledge	Comprehension	Applications	High Level Questions	Total
MEASURES OF CENTRAL TENDENCY	2	3	3	2	10
INTRODUCTION AND DATA PRESENTATION	2	3	3	2	10
MEASURES OF VARIATION	2	3	3	2	10
PROBABILITIES	2	3	3	2	10
TOTAL	10	15	15	10	40

Table 1: *Table of Specifications*

As it is noted, the above table covers the entire range of thematic sections that students are taught during Secondary School. Each section consists of two questions on Knowledge, two High Level questions (analysis, synthesis and evaluation) and three comprehension and application questions, since the subject matter of Statistics is the field of applications to a great extent.

4. Processing techniques of the study's data

The techniques used for the statistical processing of the study's data are Regression Analysis (Standardized Multiple Regression Model -SMRM), as well as, the Structural Equation Modeling (SEM), whilst the statistical software packages used were respectively SPSS (Statistical Packet for Social Sciences) and AMOS (Analysis Moment Structural).

SMRM was used because during the employment of a simple regression model, the presence of sufficient explanatory variables, as in the current study, automatically produces a multiple number of intermediate calculations, until the final results are reached, whilst approximations, even of a few significant places, especially during the inversion of the matrix $\mathbf{X'X}$ (which refers to the value of the explanatory variables $\mathbf{X}_i$), often lead to results which diverge from the correct ones [14, 15]. This effect can be avoided by using SMRM models. Furthermore, it is possible to compare the computed values of bi (regression coefficients) with those in other studies, where different measurement systems have been used. However, in the case of SMRM, the regression coefficients (bi) show the alteration that will occur at the standard deviation (s) of the Y variable from the alteration of a point at the σ_j of X_j. The relation connecting a standardized regression coefficient ($b'\kappa$) with a simple $b\kappa$ is: $b'\kappa = b\kappa/(S\psi/S\kappa)$.

Structural Equation Modeling, which forms a largely new technique, that is, a combination of Exploratory Factor Analysis and Multiple Regression Anlysis [16], offers us the option of a versatile and complete examination of the effect in question, apart from the fact that it can be used as a substantiation of the results of the Regression Analysis [17, 18]. Indeed, there are many researchers recommending the employment of the Structural Equation Modeling (SEM) technique for issues relevant to the educational research [17].

5. Application of Regression Analysis on the Sample's Data

The search for variables which influence the performance of students at the Statistics test, which they were given, is examined through a regression function of the type: $y_i = g(\mathbf{x}_i, \mathbf{b})$

Where y_i is the performance grade in the Statistics test of the i^{th} student, $\mathbf{x}_i$ is the vector of the variables, which interpret the students' performance at the Statistics test and **b** is the vector of the parameters to be evaluated. After the statistical processing of the study's data and out of the 25 candidate explanatory variables, the seven presented at tables 2 and 3 were deemed statistically significant.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	0,702(a)	0,492	0,491	1,74204	0,492	530,086	1	547	0,000
2	0,785(b)	0,616	0,615	1,51522	0,124	177,028	1	546	0,000
3	0,817(c)	0,667	0,665	1,41335	0,050	82,543	1	545	0,000
4	0,824(d)	0,679	0,677	1,38870	0,012	20,522	1	544	0,000
5	0,831(e)	0,690	0,687	1,36581	0,011	19,381	1	543	0,000
6	0,838(f)	0,702	0,698	1,34163	0,011	20,752	1	542	0,000
7	0,840(g)	0,706	0,702	1,33348	0,004	7,646	1	541	0,006

Table 2.a Model Summary

Notes:

a. Predictors: (Constant), Mathematics Performance

b. Predictors: (Constant), Performance in Mathematics, Hours Studying Statistics

c. Predictors: (Constant), Performance in Mathematics, Hours Studying Statistics, Overall Performance

d. Predictors: (Constant), Performance in Mathematics, Hours Studying Statistics, Overall Performance, Tutoring *in Statistics*

e. Predictors: (Constant), Performance in Mathematics, Hours Studying Statistics, *Overall Performance,* Tutoring *in Statistics, Parents Education*

f. Predictors: (Constant), Performance in Mathematics, Hours Studying Statistics, Overall Performance, Tutoring *in Statistics, Parents Education, Statistics Usefulness*

g. Predictors:(Constant), Performance in Mathematics, *Hours Studying Statistics, Overall Performance, Tutoring in Statistics, Parents Education, Statistics Usefulness, Students Residence*

Linear	**Explanatory Variables**	**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
Model		B	Std. Error	Beta		
	(Constant)	-4,156	0,393		-10,587	0,000
	Mathematics Performance	0,192	0,025	0,255	7,582	0,000
	Study Hours	0,534	0,058	0,301	9,258	0,000
	Overall Performance	0,226	0,033	0,226	6,898	0,000
	Tutoring in Statistics	0,319	0,064	0,122	4,981	0,000
	Parents Education	0,141	0,031	0,139	4,500	0,000
	Statistics Usefulness	0,251	0,054	0,112	4,658	0,000
	Students Residence	-0,452	0,163	-0,069	-2,765	0,006

Table2. b. *Regression Coefficients*

Note: Dependent Variable is the performance grade in the Statistics test (Statistics test).

The following regression model, which was reached

$$Y = 0{,}255X_1 + 0{,}301X_2 + 0{,}226X_3 + 0{,}122X_4 + 0{,}139X_5 + 0{,}112X_6 - 0{,}069X_7$$

The main details of the regression coefficients are represented in table 2.b above.

The check, which was performed with respect to the necessary requirements for the application of Regression Analysis had positive results.

In particular, the number of sample data was much bigger (588) than the minimum required according to international research standards -ten times the number of the explanatory variables [15]. Also, the relevant checks which were done, showed that the differences between the actual values and the values derived by the model, $\hat{y}i$ -yi = ei, are stable, as far as their variance is concerned, i.e., the Heteroscedacity effect is absent. Moreover, the e_i values follow the normal distribution. Finally, no significant degree of Multicollinearity was observed among the explanatory variables. Particularly, the inversion of the VIF (Variance Inflationary Factor) figures, produced figures close to one, which is particularly important, considering that for the possibility of Multicollinearity, the figures we produce after the inversion of the VIF must be lower than 0,2 (or 0,1 according to others) [17]. The main features of the reggression coefficients are represented on the above table 2.b.

This model interprets to a satisfactory degree the conduct (variance) of the Y variable (students' performance at the Statistics test), since the coefficient of determination was found to be equal with R^2=0,70.

As we can notice in this model, the greatest effect on the students' performance at the Statistics test is exerted by their previous year's grade in Mathematics (bi=0,255) and the number of hours allocated each week for the studying of Statistics (bi=0,301).

	Explanatory Variables	**R^2 Chang**	**-bi-**
X_1	Student's Performance in Mathematics from Last Year	0,492	0,255
X_2	Hours Studying Statistics	0,124	0,301
X_3	Student's Overall Performance from Last Year	0,050	0,226
X_4	Tutoring in Statistics	0,012	0,122
X_5	Parents' Educational Level	0,011	0,139
X_6	Statistics' Degree of Usefulness	0,011	0,112
X_7	Student's Place of Residence	0,004	0,069

Table 3. *Presentation of the* $\boldsymbol{R^2}$ *Changes, because of the gradual insertion of the explanatory variables in the Regression model.*

As we can notice from the above table -which is an excerpt of the detailed table 2.a (Model Summary)- the explanatory variable "Student's performance in

Mathematics from last year" explains the largest part (0,492) of the variance of the dependent variable Ψ, ''Student's performance at the Statistics test". The explanatory variable "Hours studying Statistics per week" is the second in degree of interpretation of the variable Ψ. The rest of the variables follow in order of their contribution to the interpretation of the variance of Ψ. As we can see, the smallest contribution to the interpretation of the variance of Ψ is attributed to the explanatory variable "Student's place of residence".

As we notice, out of the seven explanatory variables which interpret the students' performance at the Statistics test, three have social character, "tutoring in Statistics", "Parents' educational level" and "Student's place of residence", which has the least effect out of all the explanatory factors of the regression model.

6. Structural Equation Modeling (SEM)

With the aid of SEM, a versatile and in-depth analysis of the correlations between the explanatory variables and the dependent Ψ was performed, which substantiated and enhanced the Regression Analysis results [19]. With SEM, apart from the known observed variables, there are also the unobserved ones, which can be indirectly defined with the assistance of other observed variables. Moreover, with SEM one can conduct a simultaneous and complete check on all the correlations among the variables, since it provides the option for a variable which was a dependent variable in an equation, to become an independent variable in another one and vice versa [17, 20]. Also, one can create a new variable, that is called construct or latent one, with the help of another real (observed) variable that it forms the indicator variable of the latent one [20]. When we have estimated the relative model of SEM, then our aim is the population's parameters to be estimated, so that, to achieve the minimum difference between the covariances matrix of the sample and the corresponding of the estimated population. This target is achieved when the function:

$$Q = [s-\sigma(\theta)]' W [s - \sigma(\theta)] \text{ is minimum.}$$

Where: The s is the sample's covariance matrix (in the form of vector). To σ is the vector of the estimated covariance matrix for the population. The θ shows that the σ has relation with the model's parameters (regression Weights, covariances etc). To w is the matrix that weighs the differences squares between the sample covariance matrix and the corresponding of the estimated population. The w is specified, so that, the above function Q to be minimum [16].

Here, a further examination of the correlation of the variables is performed, paying particular attention to the variable "Statistics' usefulness", to the variable "Performance in Mathematics" and to the variable "Hours studying Statistics per week". This is done because "Usefulness" may be influenced by the student's particular aptitudes (e.g. in Mathematics etc.). For this reason, a SEM model is created, whereby " Statistics' Usefulness" is an endogenous variable. Also, we focus our particular interest in the study of the variable "Performance in Mathematics" which, as we have seen in Regression Analysis, out of all the variables of the regression model, along with the variable "Hours studying Statistics per week" influences to a great extent the Ψ variable (Performance at the Statistics Test). For the above mentioned reasons and for additional ones that will examine later (through the thorough study of the correlations among the variables) a SEM model is created, whereby the variables, "Hours Studying Statistics per Week", "Statistics' Usefulness" and "Student's Performance in Mathematics" along with the "Performance in Statistics" constitute the endogenous variables of the model. The variable "Hours Studying Statistics per Week", as well shall immediately examine, constitutes the indicator variable of the Latent (construct) one "Overall Study of Mathematics".

The construct (latent) variable "Overall Study of Mathematics" appears on the Model which, for brevity purposes can be simply referred to as "Study", and it expresses the degree of the student's overall study in the wider field of Mathematics (Algebra, Geometry, Statistics etc.) and which we accept that it interprets at a Level of 0,81 (r=0,9) the variable "Hours Studying Statistics per Week". The creation of the latent variable "Overall Study of Mathematics" or simply "Study" was done for the optimum interpretation of the correlation between the variable "Student's performance in Mathematics" and the variable "Hours studying Statistics per week", which constitutes the indicator variable of the latent one “Overall Study of Mathematics”. That is to say, it was assumed that a student who was generally studying Mathematics for many hours both during the previous and this year, would obviously studying Statistics for many hours this year.

The rest of the variables we examine, along with the three unobserved variables mentioned at the respective residual errors, are listed on the following table (right column) and constitute the exogenous variables cluster. We notice in the table 3, the number of the model’s variables is 13. In particular, there are eight observed variables (four endogenous variables and four exogenous ones), which are the first four in order of appearance at the two columns of table 3. Moreover, there are five unobserved exogenous variables, which appear last in the second column of table 3, four out of Which express the respective errors and the remaining one is the latent

Endogenous variables	**Exogenous variables**
Performance in Statistics	Student’s Place of Residence
Performance in Mathematics	Parent’s Educational Level
Statistics’ Usefulness	Tutoring in Statistics
Hours Studying Statistics	Student’s Overall Performance
	Error "Statistics Test"
	Error - " Statistics’ Usefulness"
	Error -"Performance in Mathematics"
	Overall Study of Mathematics
	Error -"Study"

Table 4 . The variables of the SEM model.

variable “Overall study of Mathematics”. Also, all the relevant requirements that need to be satisfied for the application of SEM are in place, so that a sound application of the model can be achieved.

The exogenous variable "Student’s overall performance" and the endogenous variable "Student’s performance in Mathematics” are exerting almost the same degree of effect to the endogenous variable “Performance at the Statistics test” (path coefficients, 0,214 and 0,204 respectively). An endogenous variable may exert influence upon another endogenous one [17]. The degree of effect exerted by the exogenous variable "Parents' educational level" and the endogenous "Statistics' usefulness", upon the variable "Student's performance at the Statistics test", is lower. The exogenous variable "Student's place of residence" exerts a low degree of influence on the endogenous variable "Students' performance at the Statistics test" (-0,097). The feature that needs to be highlighted at this point, even though the degree of this variable's effect on the students' perfomance at the Statistics test is low, is the fact that the students living in urban centers are at a more advantageous position than those in the provinces.

With respect to the effects exerted on the variable "Statistics' Usefulness" and on the "Student's Performance at Mathematics", we observe that the endogenous variables "Student's Overall Performance" and "Overall Study of Mathematics" have the greatest effects on the endogenous variable "Student's Performance at Mathematics (0,408 and 0,415 respectively). The exogenous variable "Student's Residence" has only a small effect on the

Dependent Variables	←	Independent Variables	B
Student’s Performance in Mathematics from the Last Year	←	Student’s Overall Performance from the Last Year	0,408
Student’s Performance in Mathematics from the Last Year	←	Overall Study of Mathematics	0,415
Statistics' degree of Usefulness	←	Student’s Performance in Mathematics from the Last Year	0,253
Student’s Performance in Mathematics from the Last Year	←	Student's Place of Residence	-0,089
Performance in Statistics	←	Student’s Place of Residence	-0,097
Performance in Statistics	←	Parents’ educational level	0,083
Performance in Statistics	←	Statistics' degree of Usefulness	0,112
Performance in Statistics	←	Student’s Performance in Mathematics from the last year	0,204
Performance in Statistics	←	Student’s Overall Performance from the Last Year	0,214
Performance in Statistics	←	Tutoring in Statistics	0,100
Performance in Statistics	←	Overall Study of Mathematics	0,410

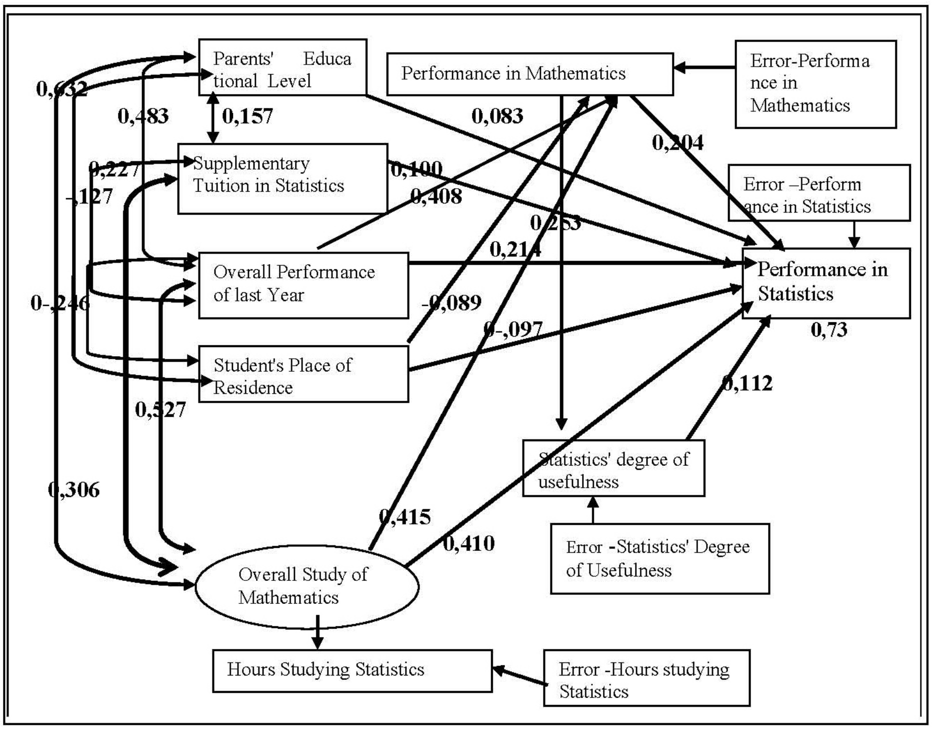
Table 5. Standardized Regression Weights

variable “Student’s Performance in Mathematics”, although the negative coefficient of this correlation (-0,089) needs to be underlined. Furthermore, the variable “Student’s performance in Mathematics” has a positive effect (0,253) on the variable “Statistics’ usefulness”. The degree of effect exerted by the latent variable ‘’Overall Study of Mathematics’’, upon the variable “Student’s performance in Statistics test”, is 0,410.

Exogenous variables		Correlations
Student's Place of Residence	Student's Overall Performance from the last year	-0,246
Student's Overall Performance from the Last Year	Overall Study of Mathematics	0,527
Parents' Educational Level	Student's place of residence	-0,127
Parents' Educational Level	Tutoring in Statistics	0,157
Parents' Educational Level	Overall Performance from the Last Year	0,483
Tutoring in Statistics	Overall Performance from the Last Year	0,227
Parents' Educational Level	Overall Study of Mathematics	0,632
Tutoring in Statistics	Overall Study of Mathematics	0,306

Table 6. Exogenous Variables Corrélations

Diagram 2. The diagram of the Structural Equation Modeling



With reference to the correlations among the model's exogenous variables, larger correlations are identified between the variables "Parents' educational level" and "Overall study of Mathematics" (0,632), as well as, between the variables "Students' Overall Performance from the Last Year" and "Overall study of Mathematics" (0,527). The correlation between the variables "Parents' educational level" and "Student's overall performance from the Last Year" (0,483) follows. Undoubtedly, in a SEM model, we are mostly interested in the correlations between the exogenous variables and the endogenous ones (Hair et al. 1999).

Finally, it is ascertained that the model applied to the sample's data interprets to a great extent the variance of the students' performance at Statistics, since R^2 =0,73. It should also be stressed that the value produced by the SEM model is close to the value provided by the typical regression analysis (R^2 = 0,70), which had initially been applied, and is slightly bigger than the latter. That is to say, it is an amplifying and confirmatory ascertainment of the initial results provided by the regression analysis.

7. The correlation between the difficulty index of the chapters of Statistics indicated by the students and the respective results they achieved at the test.

Table 4 presents the students' ratings in the form of percentages (%), so that we can know to what percentage they regard the sections of Statistics as "Easy", "Moderate", "Difficult" or "Very difficult". So, for example, we are informed from the table's first column that the students' majority, 65,8%

THEMATIC SECTION (Chapters)	Questions			
	Easy section	Moderate section	Difficult section	Very difficult section
MEASURES OF CENTRAL TENDENCY	65,8	24,1	8,7	1,4
INTRODUCTION AND DATA PRESENTATION	23,5	65,6	3,7	7,1
MEASURES OF VARIATION	3,7	7,1	71,3	17,9
PROBABILITIES	7	3,1	16,3	73,6

Table 7. *The difficulty index of the Chapters of Statistics*

of them, views "Measures of central tendency" as an "Easy Section". 23,5% of the students rates "Introduction and data presentation" as an "Easy Section", 3,7% the "Measures of variation" and finally a percentage of 7% regards "Probabilities" as an "Easy Section". The table's last column informs us that only 1,4% of Secondary

School students regards “Merasures of central tendency” as the most difficult section of Statistics, while a small percentage (7,1%) the “Introduction and data Presentation” and 17,9% the “Merasures of variation”. Apparently, the majority of the students (73,6%) rates “Probabilities” as the most difficult section of Statistics. The two columns in-between at the double entry table are interpreted in a similar manner.

The correlation coefficient of Spearman, which was used in order to calculate the exact correlation between the above estimate of Statistics’ sections difficulty index (table 4) and the respective performances achieved at the test, was found to be equal with 0,74. This is considered a satisfactory result and shows that Secondary School students, independent of their good or poor scoring at the Statistics test, are able, at least, to evaluate correctly in which sections they face difficulties.

There were no hidden surprises at the students’ performance per section. The highest scores were achieved in the section “Measures of Central Tendency” (12) and in “Organizing Data” (9,5), whereas the lowest scores were achieved in “Probabilities” (7,5) and in ‘’Measures of Variation” (8).

8. Findings.

To sum up the findings of this study, one can say that the causes (explanatory variables) that were identified through Regression Analysis, interpret to a satisfactory level the students’ performance at the Statistics test. Out of these explanatory variables, the greatest effect on the students’ performance was exerted by the “Students’ performance at Mathematics from the previous year” and the “Hours studying Statistics per week”. A moderate degree of effect is exerted on the students’ performance at the Statistics test by the variables “Tutoring in Statistics” and the “Students’ overall performance grade from the previous year”, as well as, by the “Statistics’ degree of usefulness” and the “Parents’ educational level”. Finally, of a small scale, though of a negative nature is the effect on the students’ performance at the Statistics test, exerted by the variable “Student’s place of residence”. Namely, the students of the urban areas have a little better performance on Statistics than the country students. Also, the employment of Structural Equation Modeling substantiated and enhanced the results of Regression Analysis. As we have observed, four out of the seven explanatory variables are of a social character.

Based on the above findings, we would like to suggest, that teachers, prior to teaching Statistics, conduct a diagnostic test on the areas of Mathematics that are deemed necessary for the understanding of specific Statistics units in order to cover the students' potential unfamiliarity with them. Also, we think that the use, by the teachers of various ways to attract students' interest in Statistics and demonstrate to them (to the students) its usefulness, encourage the students to spend a lot of time in the study of Statistics. The institution of supplementary tutoring within the school system and for Statistics, in particular, should be enhanced. Schools, through the implementation of a series of measures, should also (from elementary school onwards) become a cultural site, so that the influence exercised by the parents' educational level can be moderated, and the intention of studying to be stronger for all students, since the students' overall performance has an effect on their performance in the Statistics test.

Finally, the fact that the students have a sound opinion with respect to the units in Statistics that they face problems and difficulties with, is an element that helps to a certain extent towards the confrontation of the problem.

References

[1] Hubbarb, R. (1992). Teaching Statistics with minitab. *Journal citation-Australian Mathematics –Teachers, v.48, n.3*, 8-10

[2] Lossif, G. (1999). The graphics calculator as a teaching aid in statistics. *Journal citation*: *Teaching statistics v.21, n.2,* (p.p.45–48).

[3] Calzada, M. & Scariano, St. (1999). What is normal anyway? *Journal Citation: Matematics Teacher*, v.92, n.8, 682-689.

[4] Rouncefield, M. (1989). Practical Statistics in the sixth form. *Journal Citation: Mathematics in School. v.18, n.1* (p.p.6-7).

[5] Burril G. et al. (1992). Data analysis and Statistics across the Curriculum and

Evaluation Standards for School Mathematics, Addenda Series. Grades 9-12. *National Council of Mathematics. USA.*

[6] Hodgson, T. & Borkowski, J. (1998). Why stratify? *Journal citation: Teaching Statistics*, *v. 20, n.3,* 66-67.

[7] Kontogiannopoulou –Polidoridou et al. (2000), Detecting the Students' Performance in the Education of Greece. Metaichmio. Athens.

[8] Alimbinissis, A. et al. Mathématiques –3rd grade of Greek Gymnasium 1996, Athens.

[9] IASE, Training Researches in the USE of Statistics, Ed. By C. Batanero, 2001,Spain.

[10] Kassotakis, M., (1980), The valuation of the Students Performance. Athens.

[11] Godino, J. D. et al. The statistical consultancy workshop as a pedagogical. *Training researchers in the use of statistics*, p.p. 339 - 353, International Association for Statistical Education - International Statistical Institute – Tokyo, 2000.

[12] Mavromattis, G. *Statistical models and methods analysis of the data* (*p.p.11-14)*. Thessalonica: University Studio Press, 1999.

[13] Kubiszyn, T. & Borich, G. (2000). *Educational Testing and Measurement.* (6th Ed.). USA : J. Willey & Sons.

[14] Neter, j. et al. (1996). *Applied linear statistical models, edition.* (4th Ed.) U.S.A.: McGraw-Hill.

[15] Draper, N. & Smith, H. (1997), *Applied regression analysis.* Greece, Athens: Papazissis S.A.

[16] Ulman J. B. (1998), Structural Equation Modeling, in the book:*Using Multivariate Statistics*, Tabacnick et al., Wesley 3rd Ed.

[17] Hair,J. F. et al. (1998) *Multivariate Data Analysis*, N.J.: Prentice Hall.

[18] Alimissis, D. (1994). *Causes of the failure of Secondary School Students in physics*. Department: Philosophy, Pedagogical and Psychology. Unpuplished Doctoral Dissertasion, National Kapodistrian University of Athens, Greece.

[19] Douglas, T. (1983). The path analysis technique in educational research: Bridging the theory-empiricism gap. *Journal of education administration*,

v.21, (p.p. 40-51).

[20] Arbucke J. & Worthke W. (1999).*Amos 4.o User's Guide,* USA :SmallWaters corp.

◇ *Evangellos Papakonstantinou*
National Kapodistrian
University of Athens
Athens, GREECE

◇ *Dimitris Karageorgos*
National Kapodistrian
University of Athens
Panepistimiopolis15784
Athens, GREECE

◇ *Vassilios Gialamas*
National Kapodistrian
University of Athens
Athens, GREECE
gialamasbasilis@yahoo.gr

Hellenic Mathematical Society
International Journal for Mathematics in Education

Longitudinal study on mental calculation development in the first two grades of primary school

Charalambos Lemonidis

Abstract

This paper presents the results of a longitudinal experimental teaching concerning mental calculation in four basic operations for first and second graders of Primary School. Some of the characteristic features of this experimental teaching were the following: number teaching was based on a holistic and summative perspective rather than a perspective of one-to-one counting. Special emphasis was put on additive analysis and composition of numbers. Teaching was based on pupils' pre-existing knowledge and metacognitional expression of any mental strategy followed. Educational material as well as communication in the classroom was taking various semiotic representations of numbers and quantities into account. This way of teaching caused important changes in pupils' performance and attitude in general, in relation to the typical teaching, which is nowadays followed in Greece.

Keywords: Mental calculation, mental strategy, experimental teaching, flexibility in strategy, representation of quantities, metacognitive procedures

Introduction

Mental calculation is a process of calculating a result precisely without the aid of an external mean of calculation or writing. Among the calculations that adults perform in their everyday life, mental calculations are the most widely used. (Wandt, E. and Brown, G.W., 1957). Many researches point out that mental calculations play a significant role in teaching and learning mathematics. Ian Thompson (1999, p.147) underlines four basic reasons for teaching mental calculations 1) they are more commonly used than written ones; 2) they create a better and deeper understanding of number concept (Mclntosh, 1990, Sowder, 1990); 3) mental work develops problem-solving skills; 4) they help understanding and developing written methods for calculating.

There are many strategies or processes that pupils develop when working out mental calculations. In literature they can be found in categories according to the kind of operations and the size of numbers. Thus, one may find pupils' strategies for addition and subtraction of numbers up to 20 (Carpenter, T.P., Moser, J. M., 1982, Steffe, L.P., Cobb, P., 1988, K. Fuson, K.C., 1992). Strategies for addition and subtraction of numbers up to 100 (Beishuizen, 1993; Reys et al., 1995; Fuson et al., 1997). Strategies for multiplication and division (Kouba, 1989· Steffe,1994· Mulligan and Mitchelmore, 1997). etc.

There are many curricula all over the world that stress out mental calculations. For example National Numeracy Strategy that was introduced in England's primary education in 1999. Mental calculations play a vital role in N.N.S. and a structural approach to mental strategies is proposed. Direct teaching of strategies to class is also recommended. (DfEE, 1999). In the USA, National Council of Teachers in Mathematics (NCTM, 2000) is referring to pencil and paper calculations as not so important, while mental calculations, estimations and using calculators as skills that worth to be developed.

Some researches (Cooper et al., 1996a, 1996b; Heirdsfield & Cooper, 1996) mention the influence of teaching written operations to pupils' spontaneous mental strategies. Before teaching pupils present a variety of effective mental strategies, while after having being taught they tend to use a strategy that seems to reflect the written algorithm the teacher had already taught. Therefore, researchers (Kamii, Lewis, & Jones, 1991; Reys et al., 1995) recommend that pupils should be free to express their own mental strategies and that understanding algorithms improves if pupils construct strategies in accordance to their own spontaneous ways of thinking. McIntosh (1996) also supports this claim and he states that the solution to the current lack of attention in mental calculation is not to teach mental strategies in the same way that formal pencil and paper strategies had been taught in the past.

Educators view mental calculation as a mathematical process in various ways. (Reys & Barger, 1994). Reys et al., 1995, pp.304-5 distinguishes between a behavioral view and a constructivist view. A behavioral view considers mental calculation as a basic skill, perhaps serving as a prerequisite for paper/pencil computational skill or estimation, where proficiency is gained by direct teaching and practice (Shibata, 1994). According to the constructivist view, mental calculation is a higher-order thinking process where developing a strategy is as important as working it out (Resnick, 1986; Sowder, 1992).

In certain researches (Mclntosh, 1990, Sowder, 1990, 1992) the influence of learning of numbers on the ability of calculation is examined and visa-versa. However, what is not examined is the influence of number teaching on calculations considering different representations of quantities in numbers (Lemonidis, 2003). During our experimental teaching we initially presented numbers with materials (bicolour abacus and bases) that follow an organised structure using 5 and 10 as bases. In this way pupils practised in considering numbers as a sum and worked out calculations using objects.

The control group pupils were taught according to the traditional teaching that takes place nowadays in Greece. This teaching, concerning the numbers and the operations, has the following characteristics: big numbers are not taught, in the 1st grade numbers up to 20 are taught and in the 2nd grade numbers up to 100. Written operations are taught too early and no emphasis is placed on teaching mental calculation and estimation.

The hypotheses that we pose in this research are the following: a) A different presentation of numbers in a holistic and additive logic, has consequences in the way pupils work out mental calculations. The pupils reach abstract methods of calculation faster. They use a wider variety of methods in mental calculations. b) This effect of teaching is extended, apart from addition and subtraction to multiplication and division.

The main characteristics of the experimental teaching of mental calculations

The implementation of the experimental teaching on mental calculations was mainly based on the following four axes:

A) Progressive stages of the mental calculations based on the students' abilities

The students' ability to calculate in mind is a long-term target to be progressively achieved. It is well known that students at an initial level need to represent numbers with objects, in order to calculate. Our teaching aims at leading the students progressively from the concrete calculation strategy use to more abstract and mental strategy use. During the implementation of the experimental teaching the pre-existing knowledge of the students and their special needs were taken into consideration.

B) Analysis and synthesis of the numbers in a sum

The second characteristic of the teaching was the emphasis given in the analysis and synthesis of the numbers in a sum. From the beginning of the teaching, the numbers are presented analytically in the form of sums and by the help of various objects. Special focus is given to the following kind of sums:

Double sums, of the type n+n, i.e. 2+2, 3+3, etc

Analysis of the numbers in sum based either on number *"five"* or number *"ten"*.

For example, number 6 is being analyzed in 5+1, number 7 in 5+2, number 13 in 10+3 etc.

The analysis of the numbers in sums based on five and ten resembles the structure of the human fingers, and the decimal arithmetic system.

An appropriate material for their presentation apart from the fingers is the bicolor abacus (each line has ten beads, five of them are painted in different colors). Another material can be the basis of five, which consist of five cavities where children can place beads, sticks, etc. The aim is the two bases of five to be put together to form a ten.

In the beginning, the teaching was implemented with the use of small numbers and materials such as two-colored measure table and bases. Emphasis was given on the additive analysis of the numbers. In the first place, the pupils use the materials or their fingers for the calculations, and since these materials promote summing the pupils do not have to count step by step, but develop a holistic approach based on the additive analysis. We could say that the pupils are able to use *calculating strategies with the materials.*

C) Ways of semiotic representation of quantities in numbers and operations

Research has shown that the differentiation of semiotic way of presentation of a mathematics concept can also change the pupils' attitude (Duval, 1995, Lemonidis, 2003). A research Lemonidis (2003) has shown that different representations of arithmetic quantities play a very important role in teaching and learning of the first mathematical concepts. These representations can appear in different forms visual, symbolic, etc. These different expressions create different teaching situations and calculation strategies that lead to different comprehension for the students. At the first grades of primary schools, the presentation of the quantities in numbers and calculations with materials or with visual or symbolic representations can result in altering the children's attitude and that demands an alternative cognitive approach.

D) Teaching with metacognitive procedures, discussion and presentation of the "different"

In the classroom environment, the pupils' abilities differ and so do the strategies they use to implement the additions and the subtractions. Some pupils are able to use strategies of direct modeling, some others counting strategies step by step and others recall strategies (derived fact and known fact strategies)[1]. The teacher should know the strategies, which can be used by the pupils, so that he adjusts teaching to their abilities.

The teaching procedure for the calculations had the following characteristics: pupils were asked to explain every time, the method they employed for their calculations (metacognitive process). That was for the benefit of the pupils themselves, as they had to think and expose their thoughts and besides it contributed in the creation of a supportive climate in the classroom with discussions of their thoughts and calculating actions.

Research Methodology

Examination of the students' performance

Personal interviews were employed for the students' examination by a researcher who wrote a protocol for the pupil's behavior.

The whole processes lasted about 20 to 45 minutes for each pupil. All of the calculations were stated orally and the researcher noted down their oral answers on a protocol. The researcher also observed every apparent behavior, such as if and how they used their fingers, if they answered instantly or with a delay. Moreover, the researcher was trying to find out the strategies the pupil used in calculations by asking relevant questions.

[1] Some researches analyze and note the various procedures or strategies that the students use to solve simple additions and subtractions (Carpenter, T.P., Moser, J. M., 1982, Steffe, L.P., Cobb, P., 1988, Fuson, K.C., 1992).

The participants

Two groups of students participated in the research as an experimental and a control group from different schools in the city of Florina. The control group followed the traditional way of teaching and the experimental group the innovative way of teaching suggested. The participant groups were examined with seven achievement tests during two successive years in the following months: September, December, March and May of the first year and September, March and May in the second year. 14 first grade pupils participated in the first team of the control group and 15 in the second team (29 in total), whereas the experimental group included a team of 15 and a team of 16 pupils (31 in total). The same pupils in addition to some newcomers, participated the next year in class B'. So, the control group and the experimental one consisted of 35 pupils (18 pupils in one team and 17 in the other).

Table 1
Number of pupils participated in the research

1st year (class A')		**2ed year (class B')**	
Control group	Experimental group	Control group	Experimental group
14 pupils	15 pupils	18 pupils	18 pupils
15 pupils	16 pupils	17 pupils	17 pupils
Total 29 pupils	Total 31 pupils	Total 35 pupils	Total 35 pupils

The results

1. The evolution of the addition and subtraction strategies in the two samples

We will try to present an indicative picture of the gradual strategy evolution during the two-year study. We observed the strategies the two groups were using for the additions during the seven examination sessions implemented in: September, December, March and May of the first year and September, March and May in the second year. On Table 2 is presented the month and the operation taken by pupils' during their examination. The strategies were grouped into 3 major categories: recall strategies, counting strategies and direct modeling strategies.

The pupils of the two groups were examined in the beginning of the school year (September) in arithmetic concepts and were statistically found to be equal in achievement. As we can see on table 2, although the two groups started from the same point, they gradually developed different strategies.

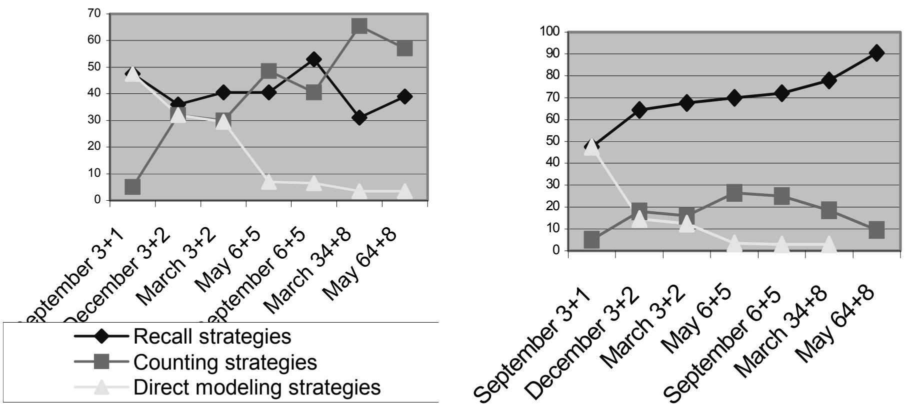
In general, we observe that the main category used by the experimental group is the recall strategies (mean 74%). They use much less the counting strategies (mean 19 %), whereas the modeling strategies are used the least (mean 6 %).

The pupils in the control group mostly use the counting strategies (45,5 %), the recall strategies (40 %) and the modeling strategies (13,5 %). In this group the counting strategies remain on top by the end of the school year in both A' and B' classes.

Table 2

Evolution of the addition strategies during two years

Control Group **Experimental group**



2. The pupils' strategies in special calculations of addition and subtraction

We will present the strategies that pupils used, in special calculations for addition and subtraction during May at the end of the first year (class A'). We selected these operations and called them *"special"*, like double sums (n+n) or the sums and complement from 10, because they are used for the working out of various other calculations. On the other hand, some of the operations like 10+n and 10n-n, show the pupils' cognition for the two digit numbers based on the numerical system. These "special" calculations are of the following:

- Double sum n+n (9+9). Complement form 2xn-n (14-7).
- One digit numbers sum equal to 10, (7+3). Difference from 10, (10-4).
- Sum form 10+n (10+6). Difference form 10n-n (16-6).
- Sum form 10n+n (14+4).

Table 3

Percentage of success and strategy use in additions and subtractions

			Recall strategies		Counting strategies		Modeling strategies	
Operation	Pupils' groups	**success**	Known fact	Derived facts	Counting without fingers	Counting fingers	Fingers	objects
9+9	Control gr	**24 (83%)**	12 (50%)	1 (4%)	1 (4%)	7 (29%)		3 (12,5%)
	Exper. gr	**29 (93,5%)**	25 (86%)	3 (10,5%)		1 (3,5%)		
14-7	Control gr	**21 (72,5%)**	3 (14,5%)	3 (14,5%)		9 (43%)		6 (28,5%)
	Exper. gr.	**29 (93,5%)**	1 (3,5%)	16 (55%)	4 (14%)	2 (7%)		6 (20,5%)
7+3	Control gr	**28 (96,5%)**	6 (21,5%)	1 (3,5%)	7 (25%)	13 (46,5%)	1 (3,5%)	
	Exper. gr.	**29 (93,5%)**	15 (51,5%)	7 (24%)	3 (10,5%)	3 (10,5%)		
10-4	Control gr	**29 (100%)**	7 (24%)	3 (10,5%)	2 (7%)	14 (48,5%)	3 (10,5%)	

	Exper. gr	**30 (97%)**	13 (43,5%)	11 (36,5%)	4 (13,5%)	1 (3,5%)		1 (3,5%)
10+6	Control gr	**29 (100%)**	16 (55%)		3 (10,5%)	7 (24%)	1 (3,5%)	1 (3,5%)
	Exper. gr	**30 (97%)**	27 (90%)		2 (6,5%)			
16-6	Control gr	**24 (83%)**	10 (41,5%)	4 (16,5%)	1 (4%)	4 (16,5%)		5 (21%)
	Exper. gr	**29 (93,5%)**	23 (79,5%)	3 (10,5%)	1 (3,5%)			2 (7%)
14+4	Control gr	**24 (83%)**	2 (8,5%)	4 (16,5%)	7 (29%)	6 (25%)	1 (4%)	4 (16,5%)
	Exper. gr.	**28 (90,5%)**	11 (39,5%)	15 (53,5%)		1 (3,5%)		1 (3,5%)

According to table 3, the pupils in the experimental group mostly use the known fact strategies (direct recall from memory): 9+9, 10+6, 16-6, 7+3 and 10-4. Double summing 9+9, of the form 10+n (10+6) and the calculation form 10n-n the majority of the pupils use the direct recall strategy from memory. For calculating the sum and complement of 10 (7+3 and 10-4), about half of the pupils use the direct recall strategy. For calculating 7+3 and 10-4, a large number of pupils (24% and 36,5% respectively) use the derived fact strategy of recollection, the *"commonly used calculations"*. For example, for the 10-4 operation, the pupils recall from memory the opposite, 6+4=10.

The pupils in the control group use the direct recall strategy in a much less degree than in the experimental group.

As a conclusion we can say that at the end of class A', the majority of the pupils in the experimental group have stored in memory and automatically recall the operations of double sums (n+n), the calculation forms 10+n and

10n+n and the sums or complement of 10.
For the calculations 14-7 and 14+4 of the form 2xn-n and 10n+n respectively, the majority of the pupils in the experimental group use the strategies of recalling commonly used operations. That is, in order to calculate the difference 14-7 they recall from memory the double sum 7+7=14 and for the calculation of 14+4 they recall 4+4; 8 and calculate 10+8=18. For these two operations the pupils in the control group mostly use counting strategies. We assume that they are not yet capable of working out these kinds of calculations, that is to recall from memory commonly used sums in order to calculate.

The flexibility in strategy use for the two pupils' groups

There was a difference in flexibility as it concerns the variety and the sort of strategies used by the two groups. The following examples are very characteristic:

In March, B' class was given the sum 8+7. The methods and operations the pupils employed for the derived fact strategy were the following:

1) Double sum 8+8 or 7+7

1.1) Double sum 8+8

8+8=16, 16-1=15 or 8+8=16, 8+7=15

1.2) Double sum 7+7

7+7=14, 14+1=15

1.3) Use of both double sums 7+7 and 8+8

7+7=14, 8+8=16, 8+7=15

2) Use of numbers over 10

2.1) 8+2=10, 10+5=15

2.2) 7+3=10, 10+5=15

3) Number analysis based on 5

3.1) 3+2=5, 5+5=10, 10+5=15

3.2) 5+5=10, 10+3=13, 13+2=15

4) Use of commonly used sums

α) 8+6=14, 8+7=15
β) 7+6=13, 13+2=15

All of the 11 pupils in the control group who used that strategy, employed double sums, 9 pupils calculated with double sum 8+8 (1.1), one pupil calculated with the double sum 7+7 (1.2) and one pupil used both of them (1.3).

The 22 pupils in the experimental group employed a wider variety of strategies:
2 pupils calculated according to 1.1
3 pupils calculated according to 1.2
10 pupils calculated according to 2.1
1 pupil calculated according to 2.2
2 pupils calculated according to 3.1
1 pupil calculated according to 3.2
2 pupils calculated according to 4, one to a) and another with b)

By the end of class B', in May, the operation 95-32 was given to the pupils. The students who used the recall strategy employed the following calculations:

1. Subtraction of the units from the units and the tens out of tens

90-30=60, 5-2=3, 60+3=63
60+30=90, 5-2=3, 60+3=63

2. Vertical subtraction

5-2=3, 9-3=6

3. First, subtraction of tens

95-30=65, 65-2=63

4. First, subtraction of units

95-2=93, 93-30=63

17 pupils in the control group correctly used the derived fact strategy and employed the following methods:

Method 1 6 pupils (35,3 %)
Method 2 10 pupils (58,8 %)
Method 3 1 pupil (5,8 %)

30 pupils in the experimental group used correctly the derived fact strategy and employed the following methods in calculating:

Method 1 16 pupils (53,3 %)
Method 2 3 pupils (10 %)
Method 3 10 pupil (33,3 %)
Method 4 1 pupil (3,3 %)

We observe from the above that the majority of the pupils in the control group employed the method of vertical subtraction for a mental calculation. On the contrary, only 10% of the pupils in the experimental used that method, whereas the majority employed methods other than the typical algorithm.

Multiplication and division

In May we conducted a research in second graders regarding their knowledge of multiplication and division

Multiplication

We proposed ten products from multiplication tables using small and big numbers as well. We tried to find representative multiplications to cover almost all the tables.

The following table shows the percentages of each strategy pupils of two groups (control group- experimental group) use. We worked on three types of strategies: Known multiplication, derived multiplication and repeated addition.

Table 4
Success percentages and strategies in multiplication

operation	Pupil s' groups	success	Known multiplication	Derived multiplication	Repeated addition
3x4	Control gr.	32 (91,5%)	19 (59,4%)	3 (9,5%)	10 (31,5%)
	Exper. gr.	32 (91,5%)	22 (68,8%)	6 (18,5%)	4 (12,5%)
3x7	Control gr.	30 (85,5%)	16 (53,3%)	7 (23,5%)	7 (23,3%)
	Exper. gr.	32 (91,5%)	19 (59,4%)	5 (16%)	7 (21,9%)
4x6	Control gr.	28 (80%)	9 (32,1%)	11 (39,5%)	8 (28,5%)
	Exper. gr.	28 (80%)	16 (57,1%)	10 (35,5%)	2 (7,1%)
5x5	Control gr.	29 (83%)	22 (75,9%)	2 (6,9%)	5 (17,5%)
	Exper. gr.	32 (91,5%)	28 (87,5%)	1 (3,1%)	2 (6,3%)
5x8	Control gr.	21 (60%)	9 (42,9%)	6 (28,5%)	6 (28,5%)
	Exper. gr.	29 (83%)	17 (58,6%)	7 (31%)	3 (10,3%)
6x7	Control gr.	14 (40%)	5 (35,7%)	6 (43%)	3 (21,5%)
	Exper. gr.	25 (71,5%)	10 (40%)	12 (48%)	3 (12%)
7x7	Control gr.	15 (43%)	7 (46,7%)	3 (20%)	5 (33,3%)
	Exper. gr.	23 (65,5%)	16 (69,6%)	7 (30,5%)	-
7x8	Control gr.	7 (20%)	1 (14,3%)	4 (57,1%)	2 (28,6%)
	Exper. gr.	20 (57%)	10 (50%)	10 (50%)	-

8x9	Control gr.	15 (43%)	7 (46,7%)	7 (46,7%)	1 (6,7%)
	Exper. gr.	24 (68,5%)	13 (54,2%)	11 (46%)	-
9x10	Control gr.	28 (80%)	28 (100%)	-	-
	Exper. gr.	34 (97%)	33 (97,1%)	-	1 (2,9%)

Experimental group pupils achieved statistically better rates than the pupils of control group in big number products, such as 5x8, 6x7, 7x7, 7x8, 8x9 and 9x10.

A first general observation is that, experimental group pupils used more developed, more abstract strategies than control group pupils. Thus, experimental group pupils used more the strategy of known multiplication and less the repeated addition strategy than the pupils did in the control group. The strategy of known multiplication, that is to say the ability of pupils to know multiplication tables by heart and to recall automatically from their memory, seems to be widely used by both groups when calculating 9x10, (100% in the control group and 97,1% in the experimental group) and 5x5 (76% and 87,5% respectively). Also, this strategy was used by the 70% of the experimental group pupils when calculating 3x4 and 7x7. Therefore we can say that the product of 10 (9x10) and the double product of 5 (5x5) have been stored in memory and are immediately used by the two groups of pupils regardless of the way of teaching. Because of the experimental teaching the particular pupils knew by heart multiplications with small numbers 3x4 and the double product 7x7. When calculating 3x7, 4x6, 5x8, 7x8 and 8x9 the experimental group pupils used the known multiplication strategy between 50% and 60%, while the control group pupils used this strategy much less (14% up to 53%.)

Only the pupils of the control group used the strategy «repeated addition with fingers or objects». This shows that these pupils fall short of the experimental group pupils regarding their ability to work out mental addition and they use objects or finger strategies in order to do repeated additions.

Divisions

The pupils were also examined in nine mental divisions (perfect divisions with a two-digit dividend and one digit divider). All divisions, apart from the 28:2 are inversely presented in multiplication tables. The performance of two groups is presented in the following table:

Table 5

Success percentages and strategies in division

			Known division	Derived multiplication		Repeated addition or subtraction
operation	Pupils' group	success		Recall inverse multiplication	Recall other operations	
28:2	Control gr.	10 (28,5%)	1 (10%)	3 (30%)	3 (30%)	2 (20%)
	Exper. gr.	22 (63%)	2 (9,1%)	12 (54,5%)	8 (36,5%)	-
12:3	Control gr.	21 (60%)	3 (15%)	13 (65%)	1 (5%)	3 (15%)
	Exper. gr.	30 (85,5%)	3 (10,3%)	20 (69%)	2 (6,9%)	4 (13,8%)
16:4	Control gr.	17 (48,5%)	1 (5,9%)	11 (64,7%)	1 (5,9%)	4 (23,5%)
	Exper. gr.	28 (80%)	1 (3,6%)	19 (67,9%)	5 (18%)	3 (10,7%)

40:5	Control gr.	13 (37%)	1 (7,7%)	9 (69,2%)	-	3 (23,1%)
	Exper. gr.	25 (71,5%)	1 (4%)	20 (80%)	4 (16%)	-
42:7	Control gr.	10 (28,5%)	-	6 (60%)	4 (40%)	-
	Exper. gr.	22 (63%)	2 (9,1%)	12 (54,5%)	6 (27,5%)	2 (9,1%)
63:7	Control gr.	8 (23%)	-	4 (50%)	4 (50%)	-
	Exper. gr.	18 (51,5%)	-	12 (66,7%)	5 (28%)	1 (5,6%)
72:8	Control gr.	12 (34,5%)	2 (16,7%)	9 (75%)	1 (8,3%)	-
	Exper. gr.	22 (63%)	3 (13,6%)	14 (63,6%)	5 (22,7%)	-
81:9	Control gr.	14 (40%)	5 (35,7%)	8 (57,1%)	1 (7,1%)	-
	Exper. gr.	22 (63%)	5 (22,7%)	14 (63,6%)	3 (13,6%)	-
80:10	Control gr.	19 (54,5%)	13 (68,4%)	4 (21,1%)	-	2 (10,5%)
	Exper. gr.	28 (80%)	13 (46,4%)	15 (53,6%)	-	-

According to the table, pupils' performance in divisions is lower than in multiplication. However the experimental group pupils' performance was much better than the control group pupils'. In all divisions, the experimental group has bigger rates of success.

The pupils' success is bigger in divisions of small numbers when the quotient is small (12:3, 16:4) and in divisions with 10 as the divider

(80:10). In these divisions about 80% of the experimental group pupils succeed, in contrast to about 50% - 60% of the control group pupils. The success of students is smaller in divisions with bigger number quotient (28:2, 40:5, 42:7, 63:7, 72:8 and 81:9). In these divisions 51,5% up to 71,5% of the experimental group pupils succeeded, whereas only 23% up to 40% of the control group pupils succeeded.

In the table above we categorized the strategies into three groups:

1) Known division, which includes the strategy of recalling a memorized division. 2) Derived multiplication, which includes two strategies: a) recall of inverse multiplication, b) recall of other multiplications or multiplications combined with additions and subtractions. 3) Repeated subtraction or addition with or without fingers or objects.

The main strategy, that pupils who divide correctly use, is recall of inverse multiplication. This means that, the success of pupils in divisions strongly depends on good knowledge of multiplication tables.

Repeated subtraction or addition with objects is almost not at all used and also repeated subtraction or addition without objects is very little used.

It is also noticeable that the experimental group pupils use more abstract strategies than the control group pupils do.

Conclusions

As we have already explained we proposed an experimental teaching, paying particular attention to the analysis and the composition of numbers in sum, to suitable semiotic presentation of numbers and to communicative learning. The results of this experimental teaching in pupils' performance were compared to those of typical teaching, which introduced written algorithms from an early stage and did not pay any attention to mental calculations.

One significant point is that, the control group pupils in the traditional teaching for addition and subtraction, by the end of second grade used mostly the counting strategies, whereas the experimental group pupils used mainly recalling and derived fact strategies. This means, that the experimental group pupils, developed and used abstract constructive strategies in all operations, earlier than the others. They had greater flexibility in using strategies of calculation, as they used a wider variety calculation than the pupils who were taught in the typical method.

We also observed that, by the end of the first grade, the experimental group pupils could understand better and calculate easier the "special additions and subtractions". These special additions and subtractions, on the one hand, are considered basic and are used for the calculation of various operations and on the other hand they show pupils' knowledge of the double-digit numbers and number system properties.

This experimental teaching, apart from the effect on addition and subtraction, had an effect on multiplication and division. We observed that experimentally taught pupils are more likely to succeed in operating multiplication and division. The experimental group pupils use more advanced strategies than the control group pupils do, such as the recall of other known operation, in order to calculate on the multiplication and the division. The flexibility of the experimental group pupils in adding has also some effect on multiplying. Thus, while some of the control group pupils use strategies with their objects or fingers in order to work out repeated additions, the experimental group pupils do not seem to use such strategies. Moreover experimental group pupils' ability to recall multiplication they knew by heart helped them to calculate better and more easily divisions as reverse operation to multiplication.

REFERENCES

Beishuizen, M. (1993). Mental strategies and materials or models for addition and subtraction up to 100 in Dutch second grades. *Journal for Research in Mathematics Education*, *24*(4), 294–323.

Carpenter, T.P. - Moser, J.M. (1982). The development of addition and substraction problem-solving skills. In Carpenter, T.P. - Moser, J.M. - Romberg, T.P. (Ed.). *Addition and Sustraction. A cognitive perspective***.** Hillsdale, Erlbaum.

Cooper, T. J., Heirdsfield, A. M., & Irons, C. J. (1996a). Children's mental strategies for addition and subtraction word problems. In: J. Mulligan, & M. Mitchelmore (Eds.), *Children's number learning* (pp. 147–162). Adelaide, SA: Australian Association of Mathematics Teachers, Inc.

Cooper, T. J., Heirdsfield, A. M., & Irons, C. J. (1996b). Years 2 and 3 children's correct-response mental strategies for addition and subtraction word problems and algorithmic exercises. In: L. Puig, & A. Guiterrez (Eds.), *International Group for the Psychology of Mathematics Education* (Vol. 20, No. 2, pp. 241–248). Valencia, Spain: University of Valencia.

DfEE (1999), *The National Numeracy Framework for Teaching Mathematics from Reception to Year 6*, DfEE, London.

Duval, R. (1995). *Sémiosis et pensée humaine: registres sémiotiques et apprentissages intellectuells.* Berne: Peter Lang.

Fuson, K. C. (1992). Research on whole number addition and subtraction. In D. Grouws (Ed), *Handbook of research on mathematics teaching and learning* (pp. 243-275). New York: Macmillan.

Fuson, K. C., Wearne, D., Hiebert, J., Murray, H., Human, P., Olivier, A., Carpenter, T., Fennema, E. (1997). Children's Conceptual Structures for Multidigit Numbers and Methods of Multidigit Addition and Subtraction. *Journal for Research in Mathematics Education*. Vol. 28, No. 2, 130-162.

Heirdsfield, A. M., & Cooper, T. J. (1996). The 'ups' and 'downs' of subtraction: young children's additive and subtractive mental strategies for solutions of subtraction word problems and algorithmic exercises. In: P. C. Clarkson (Ed.), *Technology in mathematics education. Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 261–268). Melbourne, Vic.:

Deakin University Press.

Kamii, C., Lewis, B. A., & Jones, S. (1991). Reform in primary education: a constructivist view. *Educational Horizons*, *70*(1), 19–26.

Kouba, V. L. (1989). Children's solution strategies for equivalent set multiplication and division word problems. *Journal for research in Mathematics Education*, 20, 147-158.

Lemonidis, Ch. (2003). L'enseignement des premières notions arithmétiques selon l'analyse des différentes représentations des quantités. *Annales de Didactiques et de Sciences Cognitives vol. 9, (parti 2) des actes du colloque Argentoratum 2002,* pages 103 – 117. Strasbourg, France.

National Council of Teachers of Mathematics. (2000). Principles and Standards for school mathematics. Reston, VA: The Council.

McIntosh, A. (1990). Becoming numerate: developing number sense. In S. Willis (ed.), *Being Numerate: What Counts*? pp. 24-43. Hawthorn, Victoria: ACER (Australian, Council for Educational Research).

McIntosh, A. (1996). Mental computation and number sense ofWestern Australian students. In: J. Mulligan, &M. Mitchelmore (Eds.), *Children's number learning* (pp. 259–276). Adelaide: Australian Association of Mathematics Teachers, Inc.

Mulligan, J., Mitchelmore, M., (1997). Young children's intuitive models of multiplication and division. *Journal for research in Mathematics Education*, 28, 309-330.

Resnick, L. B. (1986). The development of mathematical intuition. In M. Perlmutter (Ed.), Perspectives on intellectual development: The Minnesota Symposia on Child Psychology (Vol. 19, pp. 159 – 194). Hillsdale, NJ: Erlbaum.

Reys, B. J., & Barger, R. (1994). Mental computation: Evaluation, curriculum, and instructional issues from the United States perspective. In R. E. Reys & N. Nohda (Eds.), *Computational alternatives for the twenty-first century: Cross-cultural perspectives from Japan and the United States* (pp.31-47). Reston, VA: National Council of Teachers of Mathematics.

Reys, R. E., Reys, B. J., Nohda, N., & Emori, H. (1995). Mental computation performance and strategy use of Japanese students in grades 2, 4, 6, and 8. *Journal for Research in Mathematics Education*, *26*(4), 304–326.

Shibata, R. (1994). Computation in Japan from the Edo Era to today: Historical reflection on the teaching of mental computation. In R. E. Reys & N. Nohda (Eds.), *Computational alternatives for the twenty-first century: Cross-cultural perspectives from Japan and the United States* (pp.14-18). Reston, VA: National Council of Teachers of Mathematics.

Sowder, J. T. (1990). Mental computation and number sense. *Arithmetic Teacher*, *37*(7), 18–20.

Sowder, J. T. (1992). Estimation and related topics. In D. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 371-389). New York: Macmillan.

Steffe, L.P., Cobb, P., (1988). *Construction of arithmetical meanings and strategies*. New York: Springer-Verlag.

Steffe, L. P. (1994). Children's Multiplying Schemes. In G. Harel & J. Confrey (Eds)., *The development of multiplicative reasoning in the learning of mathematics* (pp. 3-40). NY: State University of New York Press.

Thompson, I., (1999). Written methods of calculation, in I. Thompson (ed.), *Issues in Teaching Numeracy in Primary Schools*, Open University Press, Buckingham, pp. 169–183.

Wandt, E. And Brown, G.W., (1957). Non-occupational uses of mathematics. *Arithmetic Teacher*, 4: 151-4.

Lemonidis Charalambos
Professor of Didactics of Mathematics,
University of Western Macedonia
e-mail: lemonidi@auth.gr

Hellenic Mathematical Society
International Journal for Mathematics in Education

PROSPECTIVE TEACHERS' *MORE A-MORE B* SOLUTIONS TO AREA–PERIMETER, MEDIAN-BISECTOR TASKS

Pessia Tsamir, Demetra Pitta-Pantazi

ABSTRACT

In this paper we discuss the impact of the intuitive rules on Cypriot prospective elementary school teachers' solutions to perimeter and area comparison tasks. In their solutions to the comparison-of-perimeters task, only a few prospective teachers provided a correct solution. Most of the participants were influenced by the provided drawing and answered in line with the intuitive rule more A-more B that "longer side-larger perimeter", "larger angles-larger perimeter". In their solutions to the comparison-of-areas task, about half of the participants answered correctly, but about 20% of the participants expressed more A-more B ideas.

1. INTRODUCTION

It is widely accepted that familiarity with learners' common, mathematical errors and with possible reasons for these errors should play an important role in designing instruction (e.g., Australian Education Council, 1991; Borasi, 1994; NCTM, 1991; 2000). Questions that naturally arise are: (1) How can one get to know learners' common, mathematical errors about a certain mathematical topic? and (2) How can one become familiar with possible reasons for these errors?

Knowledge about learners' common errors can be acquired mainly in two ways: (a) by reviewing the professional literature about findings of relevant studies, and (b) by carrying out related research. When looking for certain data and finding insufficient information in the literature, one obviously may design and carry out a suitable study. However, often, even when there are relevant data in the literature, researchers still find it necessary and interesting to conduct new, parallel investigations in their own country, occasionally even with their own students. In this paper we present the latter type of study. We report on a small part of an extensive study that investigated Cypriot prospective teachers' solutions to geometric comparison tasks. The formulation of this study was based on related publications (e.g., Menon, 1998; Tsamir & Mandel, 2000; Woodward & Byrd, 1983), on a matching study that was previously conducted in Israel (Tsamir, 2004), and on the aims and scope of Cypriot teacher educators.

Our study deals not only with the description of prospective teachers' correct and incorrect solutions, but also with possible reasons for their errors. While the description of learners' errors when solving mathematical tasks might be quite straightforward, the identification of the possible sources of these errors is usually much more demanding. Since understanding "why" people err is not less important than understanding "how" they err, and since theoretical models play a crucial role in such analyses, finding suitable theoretical models to account for students' solutions is highly important. In this vein, the two authors of this paper examined the applicability of the intuitive rules theory (as described, for instance, in Stavy & Tirosh, 2000; and in brief, in Section 2.1 of this paper) for analyzing the findings of both the Israeli and the Cypriot prospective teachers' solutions. We present the Israeli study in Section 2.3 of this paper and then we focus on the matching part of the Cypriot research, presenting the examination of the impact of the intuitive rules *more A-more B* and *same A-same B* on Cypriot prospective elementary school teachers' solutions to geometrical tasks. More specifically, this paper addresses the questions: (1) What are Cypriot prospective teachers' solutions to geometrical tasks, dealing with median, bisector, area and perimeter, and can these solutions be interpreted in light of the intuitive rules theory? and (2) What are similarities / differences between the findings of the Israeli study and the findings of the Cypriot study?

In the following sections we present; a theoretical background that includes a brief discussion of the intuitive rules theory, some findings about students' solutions to related geometrical tasks, and a brief report on the Israeli research (Section 2); the Cypriot study (Section 3); and a discussion of the findings with related conclusions (Section 4).

2. THEORETICAL BACKGROUND

It has been widely accepted that teachers should be familiar with (their) students' common correct and incorrect solutions to tasks embedded in various mathematical topics and with possible sources that underlie the students' difficulties; teachers should moreover use this knowledge when designing instruction (e.g., Australian Education Council, 1991; Borasi, 1986; 1987; 1992; Even & Tirosh, 2002; Fennema, Carpenter, Franke, Levi, Jacobs, & Empson, 1996; NCTM 1991; 2000; Noddings, 1992; Shulman, 1986). In this vein, when having to teach mathematics, teacher educators should be familiar with prospective teachers or teachers' typical correct and incorrect solutions, including aspects of their "knowing how" to solve mathematical tasks, and their "knowing why" certain solutions are correct and why others are incorrect (e.g., Even & Tirosh, 1995). The teacher educators should also be able to use different theoretical models that may offer an analysis of the common errors and even allow predicting various errors.

The professional mathematics education literature provides rich data regarding learners' conceptions and misconceptions concerning various mathematical notions, and there are several theoretical models that furnish different perspectives for the examination of the data (e.g., the theory of intuition in science and mathematics, in Fischbein (1987); the model of concept image and concept definition, in Tall and Vinner (1981); the model of conceptual change in science and mathematics, in Limon and Mason (2002=; and the intuitive rules theory in Stavy and Tirosh, 2000). The theoretical model we chose to use here is the intuitive rules theory. This theoretical model was chosen for several reasons. First, it has been reported to be a good explanatory framework for analyzing students' solutions to comparison tasks, like the ones used in this study. Second, the intuitive rule theory was found to be a helpful model to account for learners' various incorrect solutions and to predict and foresee students' reactions in certain settings. Finally, one of the two authors of the paper has been exploring it in various contexts, with different age groups and in a number of countries. Therefore, this study is one more piece of the longstanding investigation of the intuitive rules theory.

2.1. The Intuitive Rules Theory

Stavy and Tirosh (e.g., 2000) coined the term *intuitive rules* and formulated the intuitive rules theory for analyzing and predicting students' inappropriate responses to a wide range of mathematics and science tasks. In their examinations of students' typical responses to numerous, scientifically unrelated tasks, Stavy and Tirosh found that students tend to react to a wide range of tasks in line with three intuitive rules: *more A-more B* (e.g., Zazkis, 1999), *same A-same B* (e.g., Tirosh & Stavy, 1999) and *everything can be divided* (e.g., Stavy & Tirosh, 1993). Here, we focus on the intuitive rule *more A-more B*. Responses in line with this intuitive rule were identified in students' reactions to comparison tasks.

The intuitive rule *more A-more B* was identified in tasks in which there are two objects or systems where one quality or quantity A, fulfils the condition $A1>A2$ and this inequality is either perceptually or directly given, or alternatively, it can be logically derived through the schemes of conservation or proportion. However, participants are asked to compare the two objects or systems with regard to another quantity B, for which the two given objects or systems fulfil either $B1=B2$ or $B1<B2$. A common incorrect response to such tasks, regardless of the content domain, takes the form: “$B1>B2$ because $A1>A2$, or more A-more B”. Various *more A-more B* tendencies were identified in students' ideas regarding many mathematics and science problems (e.g., Fischbein, 1993; Noss, 1987; Stavy & Tirosh, 2000; Tall, 1981; Tsamir, 2003; Zazkis, 1999).

Another intuitive rule, *same A-same B* (or *equal A-equal B*), assists in examining inferences made by students who are asked to compare two entities with regard to a given characteristic B (where B1 is not necessarily equal to B2), based on the equality given with regard to another characteristic A (A1=A2). According to the intuitive rule *same A – same B*, the conclusion would be that if A1=A2 then B1 is also equal to B2. This rule was found to influence learners' solutions in various topics (Stavy & Tirosh, 2000; Tirosh & Stavy, 1999; Tsamir, 2002).

Although the *same A-same B* and *more A-more B* lines of reasoning are valid in some situations they do not apply to others.

Stavy and Tirosh (e.g., 2000) considered these rules *intuitive rules* since these rules bear Fischbein's (e.g., 1987) characteristics of "intuitive knowledge". That is, solutions of this type seem self evident, are used with great confidence and perseverance, and finally, they have attributes of globality and coerciveness. They further stated that their interpretation of students' incorrect responses could be regarded as a theory, which has "two main strengths: (1) It accounts for many of the observed, incorrect responses to science and mathematics tasks, and (2) it has a strong, predictive power" (Stavy and Tirosh, 2000, p. 85), making it possible to predict students' responses to certain tasks on the bases of a relevant intuitive rule.

The predictive power granted by the intuitive rules theory also served as a tool for differentiating between intuitive tasks, i.e., tasks whose correct solution is in line with an intuitive rule, and counter-intuitive tasks, i.e., tasks whose correct solution contrasts with a certain intuitive rule. In this article, we explore prospective teachers' reactions to intuitive and counter-intuitive comparison tasks. The next section surveys the literature about students' conceptions of area and volume.

2.2. STUDENTS' AREA-PERIMETER CONCEPTIONS

There are various publications reporting on students' difficulties in solving problems dealing with area and perimeter and on their tendency to regard the solutions of tasks about areas and tasks about related perimeters as being the same (e.g., Dembo, Levin, & Siegler, 1997; Hirstein, 1981; Hoffer & Hoffer, 1992; Menon, 1998; Reinke, 1997; Tsamir & Mandel, 2000; Walter, 1970; Woodward & Byrd, 1983). Students and adults were found to believe that shapes with the same perimeters must have the same area, and vice versa. They claimed that, *same area-same perimeter, same perimeter-same area.*

While most of these researchers interpreted teachers' and students' responses as resulting from a misunderstanding of the relationship between the concepts of area and perimeter, Stavy and Tirosh viewed them as instances of solutions determined by the application of the intuitive rule *same A-same B*. Moreover, Tsamir and Mandel (2000) indicated that the students' solutions in their carefully designed study, indeed exhibited strong tendencies to claim *same* area/perimeter – *same* perimeter/area, or *same* length added to two sides of a square and reduced from the other sides – *same* area/perimeter.

This article focuses on the role that a theoretical framework like the intuitive rules theory may play in analyzing prospective teachers' mathematical knowledge regarding area-perimeter median/bisector tasks. We conducted an extensive project in both countries, aimed at investigating students' and prospective teachers' intuitive solutions to mathematical tasks embedded in the related curricula. The study discussed here reports on the solutions of Cypriot prospective elementary school teachers to tasks that were formulated on the basis of a similar Israeli study, with adjustments to the Cypriot curriculum.

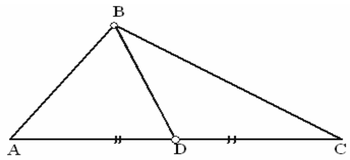
2.3. THE ISRAELI AREA-PERIMETER, MEDIAN-BISECTOR STUDY

In the Israeli study, 51 secondary school mathematics prospective teachers were asked to respond to a questionnaire (judge statements) in writing, and then, twenty prospective teachers were further asked to elaborate on their written explanations during individual interviews (Tsamir, 2004). While the participants had studied Euclidean geometry in high school and passed the matriculation examinations in mathematics upon their graduation, none of them was familiar with the intuitive rules theory.

The questionnaire included the tasks presented in Figure 1.

THE MEDIAN TASK[1]

<u>Given</u>: a. triangle ABC
b. AD = DC
c. BC > AB



Circle your solution to the following statements and explain your solution:

<u>*Statement 1*</u>

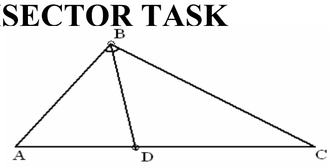
The perimeter of triangle BCD is *larger than / equal to / smaller than/ impossible to determine* **the perimeter** of ABD.

<u>*Statement 2*</u>

The area of triangle ABD is *larger than / equal to / smaller than/ impossible to determine* **the area** of BCD.

THE BISECTOR TASK

Given: a. triangle ABC
b. < ABD = < CBD
c. BC > AB



Circle your solution to the following statements and explain your solution:

<u>*Statement 3*</u>

The perimeter of triangle BCD is *larger than / equal to / smaller than/ impossible to determine* **the perimeter** of ABD.

[1] In the questionnaire "The Median Task" was called Task 7, while "The Bisector Task" appeared as Task 8.

<u>*Statement 4*</u>
The area of triangle ABD is *larger than / equal to / smaller than/ impossible to determine* **the area** of BCD.

Figure 1: The Median-Bisector Tasks in the Israeli Study

The solutions to Statements 1, 3, and 4 are that the area / perimeter of triangle BCD is larger than the area / perimeter of triangle ABD, and these solutions are in line with the intuitive rule *more A-more B* (larger side – larger area / larger perimeter). The solution to Statement 2 is that the areas are equal, and this solution is in line with the intuitive rule *same A-same B* (same lengths of sides – same areas). Since in all four statements one given, i.e., BC > AB triggers *more-more* considerations and another given (AD = DC or <ABD = <DBC) triggers *same-same* considerations, we found it interesting to examine prospective teachers' tendencies to provide erroneous intuitive solutions. Moreover, as we mentioned, three correct solutions are in line with the intuitive rule *more A-more B*, and one correct solution is in line with the intuitive rule *same A-same B;* thus, in several interviews we tried to examine whether correct solutions which were not accompanied by satisfactory explanations, were based on formal knowledge or just on intuitive ideas.

The findings indicated that, when answering the comparison tasks that addressed the perimeters, almost all the participants correctly pointed to BCD, the triangle with the larger side, as the triangle with the larger perimeter. This solution was dominant, both when referring to the perimeter-median task (94%) and when referring to the perimeter-bisector task (92%). However, only about half of the participants who correctly answered the perimeter-median task, and about a third of the participants who correctly answered the perimeter-bisector task, accompanied their solutions with a full, correct justification. While it is possible that many of the prospective teachers, who were satisfied with providing insufficient justifications, knew the comprehensive explanation, we have several examples of participants who wrote insufficient explanations to their correct judgments to the perimeter-bisector task, and indeed, based their solutions merely on intuitive reasoning. For example, during their interviews, three prospective teachers who circled the answer "the perimeter of triangle BCD is larger than the perimeter of ABD", to the perimeter-bisector task and wrote that "it's because BC is larger than AB" seemed to have based their explanation merely on *more A-more B* considerations. By the end of the interview they changed their minds regarding their correct judgment, and claimed that there were insufficient data, and thus it was impossible to determine which perimeter was larger. Following is an excerpt from Daffy's interview:

Inter.: *Here is your explanation [showing Daffy her solution] can you please explain what you meant?*
Daffy: *The perimeter [of BCD] is larger because its sides are larger…*
Inter.: *All its sides?*
Daffy: *BC is [larger]… BD is the same… so it makes no difference… and… I actually don't know about CD… I mean I dono if CD is… compared to AD…*
Interv.: *What about AD and CD?*
Daffy: *No… there is no way of knowing which one is [larger]… So, I figure that I was wrong. I can't know which perimeter is larger… I reckon that I was influenced by… that BC is larger than AB. I neglected the other sides…*

Two other prospective teachers who provided similar, written solutions ("the perimeter of triangle BCD is larger than the perimeter of ABD", because "BC is larger than AB"), did not change their minds regarding their correct judgments, but exhibited other types of erroneous, intuitive considerations. Their explanations pointed to *same A-same B* (same angle-same side) based solutions. Donna explained:

Donna: *BCD's perimeter is larger because BC is larger [than AC].*

Interv.: *And what about the other sides?*

Donna: *BC is larger than AC… BD is a side in both triangles… and… AD and CD are also equal sides… they are opposite equal angles, and in triangles opposite to equal angles there are equal sides… That's why I wrote [in her explanation in the questionnaire] only that BC is larger, because it's the only side that makes a difference.*

When replying to the area-bisector task, about 75% of the prospective teachers correctly stated that the area of triangle BCD was larger than the area of ACD, but only 10% accompanied this solution with a satisfactory explanation. The unsatisfactory explanations included (a) a partial explanation, such as, "BC is larger than AC" (about a third of the participants), (b) a vague explanation, "BC is larger than AC and the related height is also larger" with no related drawing or clarification as to why "the related height is also larger" (about 10%), and (c) no explanation.

Again, of those who provided correct judgments accompanied by insufficient justifications, several prospective teachers, in their oral interviews, came up with mere *more A-more B* considerations. Almost all the participants who did not give the correct answer that the area of triangle BCD was larger than the area of ACD claimed that they had insufficient data for answering the question.

Finally, about 60% of the prospective teachers correctly pointed to the two areas in the area-median task as being equal, about 30% accompanied this solution with a correct justification, 10% justified this judgment just by writing "AD=CD" or "the sides are equal", and the others did not explain their solution.

The incorrect solutions were that (a) "the area of triangle BCD is larger than the area of ABD" (about 20%) either because "BC is larger than AC" or because "so it seems", and (b) "it is impossible to determine which area is larger, or whether they are equal", because "there is insufficient information" (about 18%).

All in all, it seemed that the intuitive rules *more A-more B* and *same A-same B* played an important role in guiding prospective teachers' solutions to the various statements.

The following section describes the matching Cypriot study that examined prospective teachers' solutions to similar, median-bisector, area-perimeter tasks.

3. THE CYPRIOT STUDY

3.1. THE SETTING

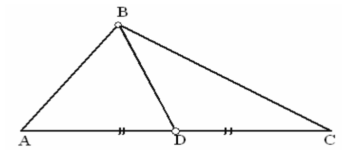
The research was conducted with 98 prospective elementary school teachers at the University of Cyprus. These prospective teachers had studied Euclidean geometry in high school and took matriculation examinations in mathematics. So, they were familiar with the theorems about triangles medians, bisectors and heights, but they had not heard about the intuitive rules theory. The second author of this paper (DPP) was teaching a Mathematics Education course in this teacher education program.

The research tools included written questionnaires and follow up, individual interviews. The questionnaires were administered during a 40-minute session of the mathematics education course. The questionnaires were administered in Greek, and they presented several problems including the ones presented in Figure 2

THE MEDIAN TASK [2]

[2] In the questionnaire "The Median Task" was called Task 1, while "The Bisector Task" appeared as Task 2.

BD is a median in a triangle ABC (that is, AD = DC). Examine the following statements:



<u>Statement 1a</u>

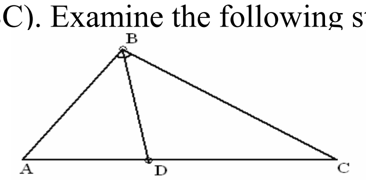
The perimeter of triangle ABD is *larger than / equal to / smaller than/ I possible to determine* **the perimeter** of BCD. (Circle your choice) Why? (Explain your answer)

<u>Statement 1b</u>

The area of triangle ABD is *larger than / equal to / smaller than/ impossible to determine* **the area** of BCD. (Circle your choice)
Why? (Explain your answer)

THE BISECTOR TASK

BD is the bisector of angle ABC in a triangle ABC (that is, angle ABD is equal to the angle DBC). Examine the following statements



<u>Statement 2a</u>

The perimeter of triangle ABD is *larger than / equal to / smaller than/ impossible to determine* **the perimeter** of BCD. (Circle your choice)
Why? (Explain your answer.)

<u>Statement 2b</u>

The area of triangle ABD is *larger than / equal to / smaller than/ impossible to determine* **the area** of BCD. (Circle your choice).
Why? (Explain your answer.)

Figure 2: The Median-Bisector Tasks in the Cypriot Study

It should be noted that the Cypriot questionnaire differed from the Israeli questionnaire in that the tasks did not include the given that BC>AC, although it seemed so in the accompanying drawing. The participants were expected to realize that the drawing does not represent a specific case, but a "general triangle". Thus, the correct solutions and the possible errors were somewhat different from the ones related to matching tasks in the Israeli study. An analysis of the solutions in line with the intuitive rules theory yields the following picture:

Statement 1a: Perimeter-Median

The correct solution – It is **impossible to determine** whether the perimeters are equal.

An incorrect solution in line with the intuitive rule *same A-same B* – The perimeters of triangle ABD and BCD **are equal**. That is, *same* lengths of segments (AD=DC) – *same* lengths of perimeters

An incorrect solution in line with the intuitive rule *more A-more B* – The perimeter of triangle BCD **is larger** than that of ABD. That is, *more* (larger) length of side (BC>BA) – *more* (larger) perimeter

Statement 1b: Area- Median

The correct solution – The areas of triangle ABD and BCD **are equal**.

An intuitive, *same-same* solution – The areas of triangle ABD and BCD **are equal**. That is, *same* lengths of segments (AD=DC) – *same* areas.

An incorrect *more-more* solution – The area of triangle BCD **is larger** than that of ABD. That is, *more* (larger) length of side (BC>BA) – *more* (larger) area

Statement 2a: Perimeter-Bisector

The correct solution – It is **impossible to determine** whether the perimeters are equal.

An incorrect *same-same* solution – The perimeters of triangle ABD and BCD **are equal**. That is, *equal* angles ($B_1=B_2$) – *equal* perimeters

An incorrect *more-more* solution – The perimeters of triangle BCD **is larger** than that of ABD. That is, *more* (larger) length of side (BC>BA) – *more* (larger) perimeter

Statement 2b: Area-Bisector

The correct solution – It is **impossible to determine** whether the areas are equal.

An incorrect *same-same* solution – The areas of triangle ABD and BCD **are equal**. That is, *equal* angles (B_1=B_2) – *same* areas.

The incorrect *more-more* solution – The area of triangle BCD **is larger** than that of ABD. That is, *more* (larger) length of side (BC>BA) – *more* (larger) area.

The analysis of the data is done with reference to this analysis.

Following the analysis of the participants' responses seven prospective teachers were individually interviewed, in order to get a better picture of their solutions and the underlying ideas. The selection of the prospective teachers was done on the basis of their responses. We chose prospective teachers who answered at least one of the questions in line with the intuitive rule *more A – more B*.

3.2. RESULTS

In this section we present the prospective teachers' tendencies to answer correctly, as well as their tendencies to answer incorrectly, with special attention to answers in line with an intuitive rule. The distributions of the participants' judgments to each task are presented in tables 1 and 2, and their related justifications are discussed and illustrated by means of relevant quotes.

3.2.1. THE MEDIAN-PERIMETER TASK

Table 1 indicates that less than 10% of the participants in this study correctly answered that it is impossible to determine which of the two triangles has a larger perimeter, or whether the perimeters are equal (in an isosceles triangle). Almost all participants (about 90%) provided the incorrect response that the perimeter of triangle BCD is larger than the perimeter of triangle ABD.

Table 1: Frequencies (%) of solutions to the **median** perimeters and areas tasks

	Perimeter (N=98)	**Area (N=98)**
Judgment		
Triangle ABD > Triangle BCD	4.1%	14.3%
Triangle ABD = Triangle BCD	-----	*49.0%
Triangle ABD < Triangle BCD	87.8%	4.1%
Undeterminable	*7.1%	31.6%
No answer	1.0%	1.0%

* Correct answer

Prospective teachers' justifications of their different responses were as follows:

Justifications of the Correct Judgment

Most of the prospective teachers who claimed that it is impossible to determine which of the triangles has a larger perimeter justified their solution by addressing the missing information regarding the lengths of sides AB and BC (5.1%). Some of them (2%) pointed to the given that AD=DC and BD=BD, and to the missing information about the relation between the lengths of AB and BC. For example,

> *It cannot be determined because we do not know the relationship of AB and BC. The remaining corresponding sides of the two triangles are equal. (Anna)*

Two prospective teachers (2%) mentioned only that AD equals DC and that needed information is missing. For example, Charalambos wrote:

> *It cannot be determined because the two triangles are not equal. That is, although they have one equal side (AD=DC) we don't know the length of the other sides.*

Penny was not completely sure about the status of the drawing, and whether the side that is drawn longer should be considered as such. She wrote:

> *[It's impossible to determine], because we don't know the length of the sides AB and BC of the triangle. If, however we assume that the length of BC is bigger than AB, then the perimeter of BCD is bigger.*

More A-more B Justifications of "BCD's perimeter is larger"

As mentioned before, most of the prospective teachers (about 90%) wrote that the perimeter of triangle BCD is larger than the perimeter of triangle ABD. Almost all of these participants (86.8%) justified their judgment in line with the intuitive rule *more A-more B*, usually by addressing the specific drawing at hand and concluding that a *larger side-larger perimeter, larger angle-larger perimeter* (84.7%), or by presenting vague consideration that "it looks larger / smaller" (2%).

The prospective teachers who elaborated on their investigations of the sides of the two triangles pointed to the two pairs of equal sides and to the last pair of sides which were unequal in the given drawing. For example, Costas, Tania and Dena wrote:

> *It depends how long AB and BC of the two triangles are because BD is common and AD=DC. Therefore we have to know AB and BC. In this case the perimeter of BDC is bigger than that of ABD because it looks like this. (Costas)*

> *One of their sides is equal and one side is common. The length of the other side for each of the triangles depends on the size of the opposite angle. Therefore the triangle's side which is opposite the bigger angle has the bigger perimeter. Therefore BCD has a bigger perimeter than ABD. (Tania).*

> *The only difference between the two triangles is the side AB (In the triangle ABD) and BC (in the triangle BCD). The triangle*

BCD is an obtuse-angle triangle whereas ABD is not ⇒ BC >AB. (Dena).

Similar justifications were voiced in the individual interviews. Here are two excerpts of Kyriako's and Chrysanthi's interviews:

Kyriakos: *It was smaller because they have AD as a common side. I considered that AD was equal to DC and BD was common. By naked eye, I considered that AB was smaller than BC and I answered smaller. When they have two elements the same and the…*

Chrysanthi: *I answered that the perimeter of triangle ABD is smaller, because AD is equal to DC. I know one side. BD is common to both triangles. That leaves AB for the specific triangle we are talking about; where visually it seems that it's smaller than BC. Therefore, the perimeter of the triangle ABD is smaller than that of BDC.*

Interv.: *When you say visually, can you explain that a little more?*

Chrysanthi: *Yes. Basically, you can understand by simply using your eyes that the length of side AB is smaller than the length of side BC.*

Another interesting phenomenon was that of complex, pseudo-formal explanations, using, for instance, the Pythagorean Theorem or differences between acute and obtuse triangles. Charalambos and Julia gave two such justifications:

AD=DC, BD common, angle BDC is obtuse; therefore the perimeter of the triangle BCD is bigger (Charalambos).

AD=DC, BD = common side, AB>BC.
$(AB)^2 = (BE)^2 + (AE)^2$

$(BC)^2 = (BE)^2 + (EC)^2$
Since AE<EC and BE common $\Rightarrow$AB<BC
$\Rightarrow$ P(ABD)<P(BCD) (Julia)

Several participants expressed the same idea by claiming that since the opposite angles to the sides are bigger then the perimeter is bigger (9.2%) or by arguing that they reached this answer based only on the visual appearance of the shape (2%). For example, George wrote: *It appears from the shape.*

All in all, when addressing the median-perimeter task, only few participants seemed to be free of the constraints imposed by the drawing, and correctly answered that the given information was insufficient for categorically pointing to one of the triangles as having a larger perimeter. In most cases, it seems that visual and *more A-more B* considerations, with no attention to the generality of the task, were dominant in the prospective teachers' judgments and in their related justifications to the median-perimeter task.

3.2.2. THE MEDIAN-AREA TASK

Table 1 indicates that about half of the participants (49%) correctly wrote that the areas of ABD and ACD are necessarily equal. About a third of the participants (31.6%) provided the incorrect judgment that it is impossible to determine which of the two triangles has a larger area, or whether their areas are equal, and about 15% of the participating prospective teachers claimed that the area of triangle ABD is larger than the area of triangle BCD.

Justifications of the Correct Judgment

The majority of the students who provided the correct response (30.6%) used the formula $S = b \cdot \frac{h}{2}$ and claimed that since both triangles ABD and

BCD have an equal sized base and the same height, their areas will also be the same. For example Panayiota drew the height BE as shown in Figure 3:

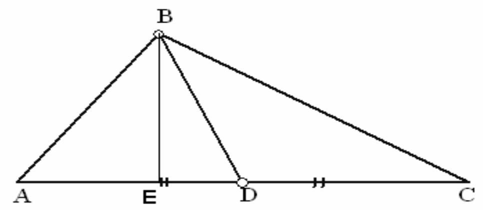


Figure 3: Panayiota's drawing

Panayiota wrote:

> *The area of ABD is equal to BCD because BE is the common height for the two triangles. Base ABD = Base BCD*
>
> *Area ABD=* $\frac{AD \cdot BE}{2}$ *Area BCD =* $\frac{DC \cdot BE}{2}$
>
> *Therefore Area ABD =Area BCD*

Other prospective teachers (16.3%) did not write the area-formula, but wrote shorter, relevant comments. For example Marianna and Zoe wrote the following

> *They [the triangles] have the same base and the same height. (Marianna)*
>
> *The height is common for the two triangles. And AD, DC are equal. (Zoe)*

Justifications of "Impossible to determine"

As mentioned before, about a third of the prospective teachers (31.6%) incorrectly wrote that it cannot be determined which of the two triangles has a larger area. Most of them (26.5%) argued that the information given in the task was insufficient. Some of them did not specify what information was missing. For example John and Elena wrote:

> *We don't have enough information. (John)*

The area cannot be determined based only on the visual observations (Elena)

Other students clearly stated that they needed some specific information. For example Nikos, Maria and Kyriakos wrote:

Because I don't know the height of the triangle. (Nikos)

We don't know how many centimeters is the base and the height of each triangle. (Maria)

We don't know either the length of the side AB or the length of the side BC. (Kyriakos)

In some instances the prospective teachers simply shared their confusion with us. Jiana and Jouly stated:

We know that the bases of the two triangles are equal but we don't know what is the height and from where to derive it. (Jiana)

We don't know which side to consider as the base of each of the triangles. (Jouly)

There were also cases where the prospective teachers considered AB and BC as the two bases of the two triangles and they drew the two respective heights, but could not reach an answer. Kate drew the following heights (Figure 4) and wrote:

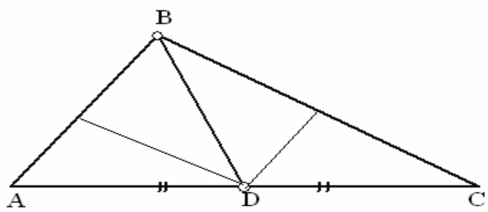


Figure 4: Kate's drawing

> *Because they have different bases and different heights and we cannot say exactly which of the two has a bigger answer in the operation* $b \times h/2$. *(Kate)*

4.1% of the participants did not explain why they provided this solution.

More A-more B Justifications of "ABD's area is larger"

Almost 15% of the participants wrote that the area of ABD was larger than that of BCE, and about 5% drew the following wrong drawing (Figure 5):

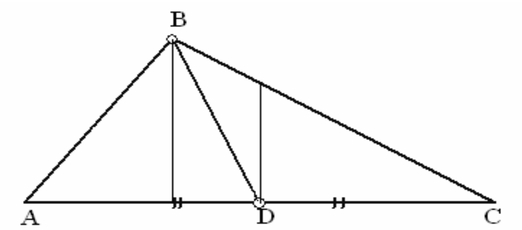


Figure 5: Christina's drawing

Christina wrote:

> *Because the area of a triangle is* $\frac{base \cdot height}{2}$. *The base (distance) is the same (AD=DC). The height of the triangle ABD, if I draw perpendiculars as shown in the drawing is bigger than the height of BCD, therefore the area of ABD is bigger.*

Two percent of the participants who claimed that the area of ABD is larger than the area of BCD drew the heights correctly but intuitively answered that "based on the fact that the height in triangle ABD is larger than the height of BCD". These students ignored the fact that the heights they drew were taken from unequal bases (AB and CB). For example Stella drew the following heights (Figure 6) and wrote:

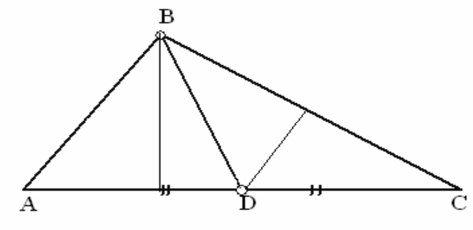


Figure 6: Stella's drawing

The height of ABD is larger than the height of BCD, therefore the area of ABD is larger. (Stella)

There was also one participant (Tasia) who claimed that the area of ABD was bigger than the area of BDC because *"it looks bigger"*.

More A-more B Justifications of "BCD's area is larger"

Several prospective teachers (4.1%) claimed that the area of triangle BCD is larger than the area of ABD. Their arguments were in line with the intuitive rule *more A – more B*. Andri and Maria drew the following heights (Figure 7) and wrote:

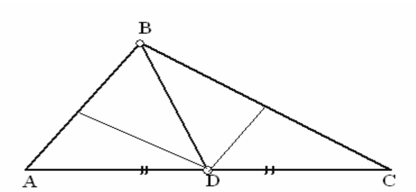


Figure 7: Maria's drawing

This can be determined based only on the observation that the base of the triangle BCD is a lot bigger, BC=6cm, than the base of the triangle ABD, where AB=3.5 cm. (Maria)

The area of triangle BCD will be larger because the base is larger and the height is still the same. (Andri)

The difference between the above two responses is that Andri visually determined the difference in the length of the sides whereas Maria measured them. Still, both of them failed to consider the fact that the bases from where they drew the heights were not equal and thus their response was not valid.

A number of participants reached their answer on the basis of the visual aspect of the shape, and wrote that "It seems like this from the shape in the drawing." One of the participants gave a response in line with the intuitive rule *same A –same B*, by arguing that the same answer should apply in the area task as it applied in the perimeter task: "For the same reason that applies for the perimeter." (John)

It should be noted that during individual interviews two prospective teachers, one who originally wrote "impossible to determine" and the other who wrote "the area of triangle ABD is larger", corrected their erroneous, written solution. Nikos, who had originally presented an "impossible to determine" judgment, and a justification similar to Kate's justification, changed his mind during the individual interview. Nikos realized that among the different, possible heights there was one preferred height that allowed him to complete the comparison.

Nikos: *I disagree with myself in the second one, but I must remember where I went wrong. I've made a mistake, I don't remember… They have common height…*

Interv.: *Which is the common height?*

Nikos: *BE. The height of ABD is BE and it's also the height of BDC. External height. Now, they have the same height, the same side, so it must be the same.*

Interv.: *Ok. So what was your mistake?*

Nikos: *Here I said that the height of this one is BE, of ABD, and the height of BDC is AZ.*

Interv.: *Isn't DZ a height of CDB? DZ?*

Nikos: *Yes.*

Interv.: *Why must you consider BE as its height? ….Why not DZ? It's also a height of the triangle.*

Nikos: *It is, but doesn't the formula say multiply the base with the height that falls on the base?*

In conclusion, when answering the median-area task, almost all of the prospective teachers who erred ignored the general perception of the task, and over-considered the given drawing in their solutions. Most of the erroneous solutions (about 32%) were that there are insufficient data provided for conclusively determining which triangle, in this specific drawing, has a larger area. Still, about 20% of the prospective teachers' justifications were in line with the intuitive rule *more A-more B* and their related judgments were either that the area of triangle ABD is larger than the area of BCD (about 14%) or the other way around (about 4%). Thus, it is reasonable to say that the *more A-more B* was quite influential on the prospective teachers' solutions to the median-area task as well.

3.2.3. The Bisector-Perimeter Task

As Table 2 illustrates, only 11.2% of the students wrote the correct judgment that *it is impossible to determine which perimeter is larger or whether the perimeters are equal*. Almost all the participants (85.7%) wrote that the perimeter of triangle BCD was larger than the perimeter of the triangle ABD. Still, few participants (3.1%) wrote that the perimeter of triangle ABD was larger than the perimeter of BCD, and accompanied the latter solution with either irrelevant justifications or no justifications at all. Following are the prospective teachers' justifications to the prevalent judgments.

Table 2: Frequencies (%) of solutions to the **bisector** perimeters and areas tasks

	Perimeter	**Area**

	(N=98)	**(N=98)**
Judgment		
Triangle ABD > Triangle BCD	3.1%	2.0%
Triangle ABD = Triangle BCD	-----	3.1%
Triangle ABD < Triangle BCD	85.7%	67.3%
Undeterminable	*11.2%	*25.5%
No answer	-----	2.0%

* Correct answer

Justifications of the Correct Judgment

Almost all of the participants (10.3%) who claimed that it is *impossible to determine* which of the two triangles has a larger perimeter justified their answer by addressing the missing information in regard to the sides of the triangles. Three of them Maria, Vasilis and Stavros claimed:

> *It cannot be determined because we do not have the information about the sides of the triangle. The angle cannot show whether a triangle has a small or big perimeter. (Maria)*
>
> *To find the perimeter of the triangles we need to have indications about the lengths of their sides. (Vasilis)*
>
> *The bisector is not the median; therefore we do not know the lengths of AD and DC. (Stavros)*

These students seemed not to be restricted by the drawing, and so they did not point to one of the triangles as the one with the larger perimeter.

More A-more B Justifications of "BCD's perimeter is larger"

As mentioned before, most of the prospective teachers (85.7%) claimed that *the perimeter of triangle BCD is larger than the perimeter of ABD.* They justified this judgment, which is in line with the intuitive rule *more A*

– *more B,* by referring to the specific drawing and concluding that *larger side – larger perimeter* (almost 62%), or by projecting a vague statement to the effect that "it looks bigger" (almost 21.7%).

Those that claimed that *larger sides–larger perimeter* based their intuitive response on the drawing. For example, Costas wrote:

> *Because all of the sides of BCD, apart from BD which is common, are bigger than the sides of ABD.*

Another interesting justification was given, for instance, by Michael who presented a triple *more A-more B, larger angle–larger side-larger perimeter,* justification:

> *The perimeter of BCD is bigger and again BC is bigger than the respective BA because it is opposite an obtuse angle. In addition to this DC is bigger than AD.*

Nana and Maria were among the prospective teachers who gave vague justifications, based on visual observations:

> *The two triangles have BD as a common side. And again it appears visually that ABD has a smaller perimeter than BDC. (Nana)*

> *It does not play any role that the angle is equal but it appears that the triangle BCD has a bigger perimeter than the triangle ABD. (Maria)*

Christina explained during her interview:

Christina: *In this exercise, we basically have the information of the bisector… from what I remember I didn't know where to use it and I used again… Ah… based on the sides, I saw it and answered smaller.*

Interv.: *Can you give a better explanation?*

Christina: Yes. Basically, we have some information concerning the angles. Based on the angles, I couldn't give an answer regarding the perimeter, so I used the sides. I saw that the sides of triangle 1 were smaller than the sides of triangle 2, I concluded that the perimeter of triangle 1 is smaller.

Interv.: When you say the sides…. Can you explain that more? That is, which did you compare?

Christina: Side BD is common. AD is smaller than DC, so is AB compared to BC.

All in all, most of the prospective teachers' solutions to the bisector-perimeter task pointed to the impact of the presented drawing and the intuitive rule *more A-more B* on their considerations.

3.2.4. THE BISECTOR-AREA TASK

Table 2 shows that only about a quarter (25.5%) of the participants provided the correct solution that *it is impossible to determine which of the two triangles has a larger area*. Most of the incorrect judgments (about two thirds) were that the area of BCD is larger than the area of ABD; very few prospective teachers wrote that the two triangles have the same area (3.1%), that the area of ABD is larger than the area of BCD (2%) or gave no answer (2%). Following are the participants' most prevalent justifications:

Justifications of the Correct Judgment

No participant provided a comprehensive justification, addressing the various types of triangles (ABD and BCD) that may be created by the bisector BD in ABC. That is, no one mentioned the options of AB=BC, AB>BC and AB<BC, which are all reasonable in this task. Prospective teachers usually justified the correct solution that *it is impossible to determine which of the two triangles has a larger area* in a vague manner, by just mentioning that the given data were insufficient for any conclusion:

It appears that the area of the triangle BDC is bigger than the area of ABD. We cannot be sure and just knowing about the two equal angles is not enough. (Marie)

We cannot determine because we do not have enough information. (Sally)

One prospective teacher, Tina, expressed an understanding that a single height can be used for calculating both the area of BCD and the area of ABD. She drew the height (Figure 8) and posed some question marks regarding the sizes of the related bases:

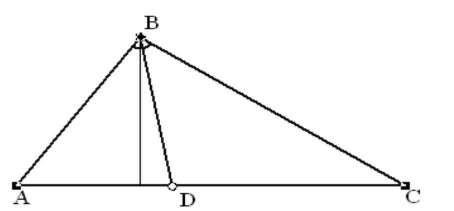


Figure 8: Tina's drawing

Because we do not know how long AD and DC are.

A small number of prospective teachers addressed the specific drawing as a given, tried to examine one set of heights and related bases in this drawing, and then pointed to an exactly defined, missing piece of information. For example, Evis addressed the drawing, added two heights to it, considered this to be the only possibility to draw heights in this task, (Figure 9) and wrote about his dilemma – the base of triangle BCD is larger, while its height is smaller, so how can one know whether its area is larger or smaller?

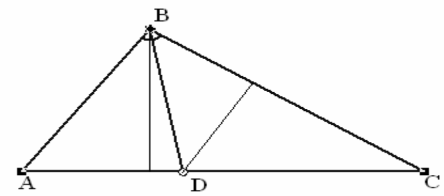


Figure 9: Evi's drawing

Because whereas the height of BDC is smaller than the height of ABD, at the same time the side DC is bigger than AD, and therefore we cannot determine with accuracy...

Dora expressed a similar idea with no explicit reference to the drawing, and Kerri was bothered only by the absence of data regarding the heights:

Basic information is missing, such as how many cm is the base and height of the triangle. (Dora)

We cannot determine because we do not know the height of the triangles. (Kerri)

More A-more B Justifications of "BCD's area is larger"

Most of the prospective teachers (67.3%) erroneously answered that *the Area of BCD is larger than the area of ABD,* and most of the justifications for this erroneous solution were based on the given drawing, expressing *larger base – larger area, larger side – larger area, larger height – larger area* ideas, in line with the intuitive rule *more A –more B.* A few participants provided vague comments to the effect that "it looks larger".

Participants who addressed the specific characteristics of the provided illustration (44.6%) argued that since the heights of the two triangles were the same, the larger basis yielded a larger area. Theo and Rea wrote:

The height remains the same but the base is larger (Area $= \frac{base \cdot height}{2}$ *) (Theo)*

Area $= \frac{base \cdot height}{2}$*. Their height is the same (BE). The base of BCD is bigger* $\Rightarrow$ *the area of ABD<BCD (Rea)*

Several prospective teachers (3.1%) drew the following heights (Figure 10) and gave a response similar to the one given by Maria:

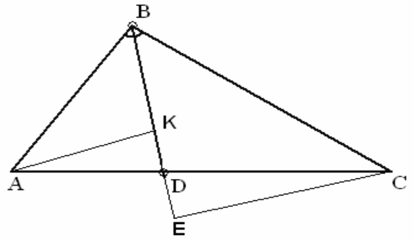


Figure 10: Maria's drawing

> *Area(ABD)=* $\frac{AK \cdot BD}{2}$, *Area(BCD)=* $\frac{EC \cdot BD}{2}$. *Although we cannot determine this with a proof, we know that BD is common and from the appearance of AK<EC I conclude that Area (BCD) > Area (ABD).*

Several justifications (about 4%), which were also in line to the intuitive rule *more A – more B* offered *larger perimeter – bigger area* links between the previously solved perimeter-task and the present area-task. For example, Anna and Renos wrote:

> *The triangle ABD has smaller sides therefore it has a smaller area. (Anna)*

> *Since the perimeter is smaller, therefore the area is also so (the quantities are proportional). (Renos)*

Prospective teachers, who accompanied their "the area of BCD is larger" judgments with vague comments (12.5%), presented justifications similar to the ones given by Marina and Nikos:

> *It appears like this from the shape. (Marina)*

We don't know either the height or the base of the two triangles, but visually the triangle BDC has a bigger area. (Nikos)

Same A-same B Justifications of "the areas are equal"

Only 3.1% of the participants judged the *area of ABD as equal to the area of BCD.* Their justifications were in line with the intuitive rule *same A – same B*. For example John wrote:

It is the same because the heights and the bases are common, therefore if we apply the formula $S=b\cdot\frac{h}{2}$ *we will see that it is the same. (John)*

More A-more B Justifications of "ABD's area is larger"

Only two prospective teachers (2%) judged ABD's area to be larger. While both provided a *greater height – larger area* justification that ignored inequality of the relevant bases and was in line with the intuitive rule *more A-more B,* one of them accompanied her justifications with the following drawing (Figure 11):

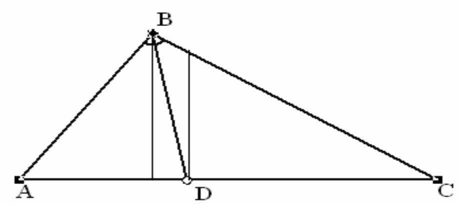


Figure 11: Penny's drawing

The area of ABD is larger because its height is larger. (Penny)

In sum, most of the participants' responses to the bisector-area task were erroneous and seemed to be influenced by intuitive, *more A-more B* or *same A-same B* considerations.

4. Some Concluding Comments

In the introduction we posed two questions: (1) what are Cypriot prospective teachers' solutions to geometrical tasks, dealing with median, bisector, area and perimeter, and can these solutions be interpreted in light of the intuitive rules theory? And (2) what are the similarities / differences between the findings of the Israeli study and the findings of the Cypriot study? The findings will be discussed with reference to these questions, and will be followed by some suggestions for related educational implications.

4.1. Cypriot Prospective Teachers' Solutions and their Interpretation in Light of the Intuitive Rules Theory

This paper describes Cypriot prospective teachers' solutions to two comparison-of-perimeters tasks and two comparison-of-areas tasks. One pair of perimeter-area tasks related to triangles created by a median, and another pair of perimeter-area tasks related to triangles created by a bisector. While the correct judgment of the statement regarding the median-area task is that the areas are equal, the correct judgment of the other three statements is that it is impossible to determine which of the two triangles has a larger perimeter / area or whether the perimeters/ areas are equal. Our findings indicated that the highest percentage (about 50%) of correct judgments and correct justifications was found in the prospective teachers' solutions to the median-area task, where the answer "equal areas" is suitable both in this specific case as well as in general. However, about 20% of the participants argued either that triangle BCD had a larger area than triangle ABD or that triangle ABD was the one with the larger area, and based their explanations on *more A-more B* considerations.

In the tasks whose correct solution is that it is impossible to determine which triangle has a larger perimeter / area, most participants erroneously pointed to one of the triangles as having a larger perimeter / area, and their justifications were based on visual considerations related to the given drawing and on intuitive, *more A-more B* ideas. More specifically, when comparing the perimeters both in the median-perimeter task and in the bisector-perimeter task over 85% of the participants answered that the perimeter of triangle BCD was larger than the perimeter of triangle ABD. Their justifications were of the type "longer side-larger perimeter", "larger angles-larger perimeter", i.e., in line with the intuitive rule *more A-more B.* A considerable, but smaller number of participants (about 69%) exhibited considerations in line with the intuitive rule *more A-more B* when pointing either to BCD or to ABD as the triangle with the larger area in the bisector-area task. Here, we also found several prospective teachers (about 3%) who used intuitive, *same A-same B* ideas.

Consequently, it is clear that prospective teachers' solutions could be interpreted in light of the intuitive rules theory, and many of these solutions were found to be consistent with the intuitive, *more A-more B* or *same A-same B* lines of reasoning (Stavy & Tirosh, 2000; Tirosh & Stavy, 1999). In most cases the participants based their arguments on their visual grasp of the data in the illustration and quite frequently they accompanied their *more-mor*e claims with pseudo-formal proofs (Vinner, 1997).

4.2. Similarities / Differences Between the Findings of the Israeli Study and the Findings of the Cypriot Study

This paper presents new data regarding Israeli and Cypriot prospective teachers' tendencies to provide erroneous, intuitive rules based solutions to similar comparisons tasks. Before going into the discussion of the similarities and the differences between the findings of the two studies, we would like to point to the importance that we see in such similar or replica studies carried out with parallel or varied populations in different countries. The importance of such studies lies in two aspects. On the one hand, they provide extended data regarding learners' ways of thinking about *a specific mathematical topic*. In this way the mathematics education community gets a better picture, for example, regarding intuitive pitfalls hidden in this topic and regarding possible reasons for students' difficulties. On the other hand, replica studies provide an extended basis for the interpretation of learners' *mathematical reasoning in the wide sense*, giving more weight to the predictive power of related theoretical models.

Here, the formulation of the tasks in the Israeli study was based on the intuitive rules theory, and the formulation of the tasks in the Cypriot study was done with reference to the Israeli study. Each of the researchers is planning instructional tools for teacher education programs in her institute, and consequently, these findings will serve in their related design. Moreover, the findings strongly substantiate the impact of the intuitive rules theory of learners' geometrical solutions, providing more evidence to the validity and the usefulness of this theoretical model.

The two studies yielded similar frequencies of the prevalent solutions to all tasks. That is, we found similar tendencies to provide "equal areas" judgments to the median-area tasks (60% of the Israeli participants and about 50% of the Cypriot participants), similar frequencies of "unequal areas" judgments to the bisector-area tasks (75% of the Israeli participants and about 67% of the Cypriot participants), and similar frequencies of "unequal perimeters" judgments to both the median-perimeter tasks (94% of the Israeli participants and about 88% of the Cypriot participants) as well as to the bisector-perimeter tasks (92% of the Israeli participants and about 86% of the Cypriot participants).

While at first glimpse, this similarity triggers one to conclude that in both countries prospective teachers had similar response patterns, a further look shows that this is not precisely the case. Due to the extra given in the Israeli study (BC>AC) the Israeli "unequal" judgments to the three latter tasks were correct, while the Cypriot ones were incorrect (here the correct judgment was "impossible to determine"). Should we conclude that the Israeli prospective teachers exhibited better knowledge? Our assumption is that this is not the case. We believe that both Israeli and Cypriot participants have a strong tendency to consider the given drawing in their solution, and consequently, to limit the generality of the given task. Here, due to the different phrasing of the statements, this tendency was acceptable in the Israeli study but unacceptable in the Cypriot study. All in all, in both studies we found, participants' justifications to be in line with the intuitive rule *more A-more B*. In the Cypriot study these were evident both in the written justifications as well as in the oral interviews, and in the Israeli study such evidence were found especially during the individual interviews. This finding should be further investigated.

Most interesting were findings regarding the "unequal areas" judgments to the median-area tasks. While in both countries about 20% of the participating prospective teachers wrote "unequal areas" answers, the Cypriot participants (about 14%) thought ABD had a larger area, and the Israeli participants regarded BCD as having a larger area. Both populations provided justifications in line with the intuitive rule *more A-more B*. The Cypriot prospective teachers usually added inappropriate heights to the given drawing, while the Israeli prospective teachers were usually satisfied with merely pointing to the given BC>AC, which is irrelevant in this statement. The two studies provide a rich scope of intuitive, *more A-more B* and *same A-same B* solutions, and indicate, like previous socio-cultural studies, that the intuitive rules influence learners' reasoning in various countries (e.g., Stavy, Tsamir, Tirosh, Lin, & McRobbie, 2001; Tsamir, Lin, & Stavy, 2001; Zazkis, 1999).

Still, there is a need to further examine the impact of the intuitive rules on learners' solutions in various socio-cultural frameworks: Do the intuitive rules influence learners' solutions to various tasks in other countries too? If the answer to this question is positive, what solutions does this influence yield? And how can this knowledge be used when designing instruction? These questions should be examined by parallel research studies in various countries.

4.3. Some Educational Implications

While there is a wide consensus in the professional literature regarding the need to consider students' ways of thinking, and their common errors when designing instruction, there are no conclusive suggestions regarding "how to take these considerations" from theory to practice. Different researchers (e.g., Fischbein, 1987; Stavy & Tirosh, 2000; Vosniadou, Ioannides, Dimitrakopoulou, & Papademitriou, 2001) agree that it is important to promote the learners' awareness of their correct and incorrect ways of thinking, and of possible reasons for their errors. When the subjects are teachers or prospective teachers (as in both, the Israeli and the Cypriot study) we believe that it is important to promote the participants' awareness of (a) the vulnerability of their geometrical knowledge, where and why they err, and (b) the power of the intuitive rules theory as an explanatory and predictive tool for addressing learners' mathematical solutions. This awareness may contribute both to their Subject mater knowledge and to their Pedagogical content knowledge (Shulman, 1986).

A number of studies have indicated that prospective teachers' familiarity with the intuitive rules theory is most beneficial. Many of the prospective teachers who had studied this theory promoted their mathematical knowledge of topics that they were about to teach, and enhanced their related pedagogical content knowledge. Beyond the acquisition of knowledge, the intuitive rules theory served these prospective teachers in refining their "listening skills" and their *alertness* to students' erroneous, intuitive-rules based solutions. Research findings indicate that prospective teachers who were familiar with the intuitive rule theory were able to identify intuitive rule based error made in class both when observing lessons and when teaching. Consequently, familiarity with the intuitive rules theory contributed to the prospective teachers' knowledge and skills which are regarded essential for teaching (e.g., Tsamir, 2006; 2007; in press).

Moreover, it seems that their familiarity with the intuitive rules theory may grant prospective teachers with tools for interpreting the reasoning they had witnessed in class, and give them professional confidence to decide that something should be done. This professional confidence, founded on knowledge, is valuable, since commonly novice teachers are preoccupied with "survival" issues, which leave them with limited energies for didactics (e.g., Borko, Eisenhart, Brown, Underhill, Jones, & Agard, 1992).

Clearly, questions like, what are the benefits of presenting other theoretical models in teacher education programs? how can teacher educators' familiarity with theoretical models, like the intuitive rules theory serve in promoting prospective teachers' SMK and PCK? and in what ways should teachers be exposed to the intuitive rules theory? should be further investigated.

For this purpose we intend to discuss with Israeli and Cypriot prospective teachers the intuitive rules theory, ask them to solve tasks like the ones presented here and then to react to various solutions given by other participants (like the solutions presented here). Studies have shown that a profound mathematical knowledge may assist in controlling intuitive ideas (e.g., Fischbein, 1987). In this spirit, we will ask the prospective teachers, what, in their opinion, are the mathematical notions that needed further clarification in the given solutions? We believe that the notions "height" in a triangle, areas of triangles, and relationships between equal angles and equal sides should be discussed. Prospective teachers will also be asked to present these tasks to students in classes that had studied the relevant notions and examine the students' solutions. Clearly, these instructional steps should be carefully designed in details, implemented, and their impact on learners' mathematical performance should be examined. More generally, the question: how can teacher educators' familiarity with theoretical models, like the intuitive rules theory serve in promoting prospective teachers' SMK and PCK? should be further investigated.

REFERENCES

Australian Education Council (1991). *A national statement on mathematics for Australian schools*. Melbourne: Curriculum Corporation.

Borasi, R. (1986). Using errors as springboards for the learning of mathematics. *For the Learning of Mathematics, 7*, 1-13.

Borasi, R. (1987). *Exploring mathematics through the analysis of errors.* Montreal, Quebec, Canada: FLM Publishing Association.

Borasi, R. (1992). *Learning mathematics through inquiry*. Portsmouth, NH: Heinemann Educational Books.

Borasi, R. (1994). Capitalizing on errors as "springboards for inquiry": A teaching experiment. *Journal for Research in Mathematics Education*, *25*, 166-208.

Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education, 23*, 194-222.

Dembo, Y., Levin, I., & Siegler, R. S. (1997). A comparison of the geometric reasoning of students attending Israeli ultra-orthodox and mainstream schools. *Developmental Psychology, 33,* 92-103.

Even, R., & Tirosh, D. (1995). Subject-matter knowledge and knowledge about students as sources of teacher presentations of the subject-matter. *Educational Studies in Mathematics, 29*, 1-20.

Even, R., & Tirosh, D. (2002). Teacher knowledge and understanding of students' mathematical learning. *Handbook of international research in mathematics education* (pp. 219-240). Erlbaum.

Fennema, E., Carpenter, T., Franke, M., Levi, L., Jacobs, V., & Empson, S. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, *27,* 403-434.

Fischbein, E. (1987). *Intuition in science and mathematics: An educational approach.* Dordrecht, Netherlands: Reidel.

Fischbein, E. (1993). The interaction between the formal, the algorithmic and the intuitive components in a mathematical activity. In R. Biehler, R. W. Scholz, R. Straser, & B. Winkelmann (Eds.), *Didactics of mathematics as a scientific discipline* (pp. 231-245). Dordrecht, Netherlands: Kluwer.

Hirstein, J. (1981). The second area assessment in mathematics: Area and volume, *Mathematics Teacher, 74,* 704-708.

Hoffer A.R., & Hoffer, S.A.K. (1992). Geometry and visual thinking. In T.R. Post, (Ed.), *Teaching mathematics in grades k-8: Research-based methods* (2nd Ed), Boston: Allyn and Bacon.

Limon, M., & Mason, L. (Eds.) (2002). *Reconsidering conceptual change: Issues in theory and practice.* Dordrecht, Netherlands: Kluwer.

Menon, R. (1998). Preservice teachers' understanding of perimeter and area. *School Science and Mathematics, 98*(7), 361-368.

National Council of Teachers of Mathematics [NCTM]. (1991). *Professional standards for teaching mathematics*, Reston, VA: NCTM.

National Council of Teachers of Mathematics [NCTM]. (2000). *Principles and Standards for School Mathematics*, Reston, VA: NCTM.

Noddings, N. (1992). Professionalization and mathematics teaching. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 197-208). New York: Macmillan.

Noss, R. (1987). Children's learning of geometrical concepts through LOGO. *Journal for Research in Mathematics Education, 18*, 334-362.

Reinke, K. S. (1997). Area and perimeter: Prospective teachers' confusion. *School Science and Mathematics, 97*(2), 75-77.

Shulman, L. S. (1986). Those who understand: Knowledge growth with teaching. *Educational Researcher,* 15 (2), 4-14.

Stavy, R., & Tirosh, D. (1993). Subdivision processes in mathematics and science. *Journal of Research in Science Teaching, 30,* 579-586.

Stavy, R., & Tirosh, D. (2000). *How students' (mis-)understand science and mathematics: Intuitive rules*. New York: Teachers College Press.

Stavy, R., Tsamir, P., Tirosh, D., Lin, F. L., & McRobbie, C. (2001). Intuitive rules: A cross-cultural study. *Paper presented at the 3rd International Conference of the European Science Education Research Association (ESERA 3)*. Thessaloniki: Greece.

Tall, D. (1981). Intuitions of infinity. *Mathematics in School*, *10*, 30-33,

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics, with special reference to limits and continuity. *Educational Studies in Mathematics, 12,* 151-169.

Tirosh, D., & Stavy, R. (1999). Intuitive rules: A way to explain and predict students' reasoning. *Educational Studies in Mathematics, 38,* 51-66.

Tsamir, P. (2002). The intuitive rule same A – same B: The case of triangles and quadrilaterals. *Focus on Learning Problems in Mathematics, 24,* 54-70.

Tsamir, P. (2003). Different representations in instruction of vertical angles. *Focus on Learning Problems in Mathematics, 25*, 1-13.

Tsamir, P. (2004). Students' intuitive rules based solutions: The case of median, bisector, area and perimeter. Unpublished research report [in Hebrew].

Tsamir, P. (2006). Enhancing prospective teachers' knowledge of learners' "same A--same B" conceptions: The case of area and volume. *Journal of Mathematics Teacher Education. 8*, 469-497.

Tsamir, P. (2007). When intuition beats logic: Prospective teachers' awareness of their same sides -same angles solutions. *Educational Studies in Mathematics 65*(3), 255-279.

Tsamir, P. (in press). Using theories as tools in teacher education. In T. Wood (Series Ed.) & D. Tirosh (Vol. Ed.), *International handbook of mathematics teacher education: Vol. III. Tools and Processes in Mathematics Teacher Education*. Rotterdam, the Netherlands: Sense Publishers.

Tsamir, P., & Mandel, N. (2000). The intuitive rule same A – same B: The case of area and perimeter. In T. Nakahara, & M. Koyama (Eds.), *Proceedings of the 24th Annual Meeting for the Psychology of Mathematics Education* (Vol IV, pp. 225-232). Hiroshima: Japan

Tsamir, P., Lin, F. L., & Stavy, R (2001). Investigating the Influence of the intuitive rules in Israel and in Taiwan. In M. van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th Annual Conference for the Psychology of Mathematics Education* (Vol IV, pp. 311-318). Utrecht: Holland.

Vinner, S. (1997). The pseudo-conceptual and the pseudo-analytical thought processes in mathematics learning. *Educational Studies in Mathematics, 34,* 97-129

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E., 2001, Designing learning environments to promote conceptual change in science. *Learning and Instruction*, *11*, 381-419

Walter, N. (1970). A common misconception about area. *Arithmetic Teacher, 17,* 286-289.

Woodward, E., & Byrd, F. (1983). Area: Included topic, neglected concept. *School Science and Mathematics, 83,* 343-347.

Zazkis, R. (1999). Intuitive rules in number theory: Example of 'The more of A, the more of B' rule implementation. *Educational Studies in Mathematics, 40*(2), 197-209.

Pessia Tsamir
Department of Science Education
School of Education
Tel Aviv University
Ramat Aviv 60078
ISRAEL
pessia@post.tau.ac.il

Demetra Pitta-Pantazi
Department of Education
University of Cyprus
P.O. Box 20537
1678 Nikosia
CYPRUS
dpitta@ucy.ac.cy

Hellenic Mathematical Society
International Journal for Mathematics in Education

Pratiques de l'évaluation formative des apprentissages: une étude de cas multiples menée auprès d'enseignants des mathématiques du secondaire

Corneille Kazadi, Ph.D.

RÉSUMÉ

Cherchant à comprendre les difficultés des enseignants[1] dans la mise en pratique de l'évaluation formative et à réfléchir sur l'écart qui existe entre le discours et les pratiques concernant celle-ci, notre recherche s'est intéressée aux pratiques des enseignants de mathématiques du secondaire à l'égard de l'évaluation formative, et aux principes qui guident ces enseignants dans l'action. S'inscrivant dans un paradigme interprétatif, cette étude vise donc avant tout à mieux comprendre le sens que l'acteur, ici l'enseignant, donne à l'évaluation formative en contexte. Une étude de cas multiples a été menée à cette fin auprès de cinq enseignants de mathématiques du secondaire. Des données provenant, pour chacun des ces enseignants, d'observations en classe et d'entrevues individuelles ont été analysées. Les résultats permettent d'expliciter la manière dont ces enseignants donnent sens au quotidien de la classe à l'évaluation formative. Une diversité de pratiques évaluatives formatives prennent place en classe, dans l'informel autant que dans le formel, dépassant la simple utilisation du test écrit instrumenté.

Mots clés de l'article : didactique des mathématiques, pratiques évaluatives, évaluation formative.

ABSTRACT

Starting from the difficulties that teachers have in carrying out formative assessment and from the gap that exists between discourse and practices about formative assessment, our research tries to better understand practices of secondary mathematics teachers in regard to formative assessment. This research wants also to put in light the underlying principles that guide teachers in those practices. Inscribed in an interpretative paradigm, this research tries in fact to better understand the meaning that the actor, in that case the teacher, gives to formative assessment in context. A multiple cases study was conducted involving five high school teachers in mathematics (secondary 1 to 5). Data coming from observations in class and interviews with each teacher were analyzed. Our results explicit the way five teachers who had participated to this study used formative evaluation in the classroom. A diversity of practices takes place in an informal as well as in a formal way, going beyond written instrumented tests.

Keyswords: didactic of mathematics, evaluative practices, formative evaluation.

Introduction

Depuis quelques années, l'évaluation occupe une place importante dans les systèmes scolaires dans le monde et de plus en plus dans le système scolaire québécois, notamment au secondaire, en mathématiques. Les évaluations fréquentes, provenant du ministère de l'Éducation du Loisir et du Sport (MELS), des commissions scolaires ou du milieu local ont en effet pris une place de plus en plus grande au sein de l'école, cherchant à cadrer la réussite des élèves et à contrôler la performance du système scolaire.

La préoccupation d'optimisation de la réussite scolaire et la valorisation du cheminement personnel de l'élève dans son apprentissage ont amené à des formes d'évaluation, dites formatives, centrées sur la prise en compte des difficultés individuelles des élèves dans l'apprentissage de concepts en mathématiques (Conseil Supérieur de l'éducation, 1987, 1992; MEQ, 1993; Bednarz et Garnier, 1989). De Cotret et Dassa (1989) soulignent que ces évaluations cherchent à repérer les erreurs des élèves, à décrire leur nature, le contexte pédagogique où elles se produisent et tentent d'en analyser les causes possibles en regard de l'apprentissage de l'élève ou de l'enseignement reçu préalablement, proposant des interventions correctrices. Ces évaluations renseignent sur le processus d'apprentissage des élèves et aident à cerner des interventions qui prennent en compte ces erreurs.

Or ce repérage des erreurs des élèves par les enseignants dans ces évaluations et l'analyse de productions d'élèves n'est pas toujours quelque chose de simple pour l'enseignant. Ça exige de plus beaucoup sur le plan de la gestion en classe de celles-ci, pour prendre en compte l'ensemble des élèves, la construction de moyens d'évaluation, l'analyse des productions d'élèves, dans le suivi auprès des élèves, etc. Nous nous sommes intéressés à cette question de l'évaluation formative en contexte, pour mieux comprendre au départ la forme qu'elle revêt dans la pratique et les défis qu'elle pose à l'enseignant.

1. Problématique, objectifs et cadre conceptuel de la recherche

1.1 Problématique

L'évaluation a toujours constitué une partie importante du travail de l'enseignant. Dans le domaine spécifique de l'enseignement des mathématiques, les faits d'évaluation occupent une place centrale au regard notamment des orientations mises en place par le curriculum au secondaire (Programmes du M.E.Q. de 1993 et de 2003) en s'assurant que les pratiques d'évaluation soient de plus en plus liées aux apprentissages essentiels proposés dans les programmes d'études.

Toutefois, les quelques recherches réalisées auprès d'enseignants en mathématiques (Groupe didactique, 1999 ; Van Nieuwenhoven et Jonnaert, 1992) ou d'enseignants de différents niveaux et différentes matières d'une façon générale (Conseil Supérieur de l'Éducation, 1987, 1992 ; Dassa et Dumoulin, 1991; Dumoulin, 1991; Forgette-Giroux, Bercier-Larivière et Simon, 1996 ; McMorris et Boothroyd, 1993 ; Parent, Seguin, Gadbois et Burelle, 1993 ; Scallon, 2000) laissent penser que cette évaluation, sous sa fonction formative, réside davantage au niveau des projets qu'au niveau des actions et des pratiques réelles.

Ces études soulignent qu'au niveau des actions et des pratiques réelles plusieurs enseignants éprouvent encore une certaine confusion au niveau des concepts de base utilisés en évaluation formative des apprentissages. Ainsi dans le Rapport du Conseil Supérieur de l'Éducation (1987, 1992), l'évaluation des apprentissages au secondaire apparaît dans une « zone grise » et là où elle se pratique, l'évaluation formative sert à l'évaluation sommative (elle prépare en quelque sorte l'évaluation sommative). Le Conseil Supérieur de l'Éducation (1992) note par ailleurs que les enseignants perçoivent généralement l'évaluation comme étant un processus distinct de celui de l'enseignement, ils négligent la rétroaction et réinvestissent peu les résultats d'évaluation dans le processus d'apprentissage.

Ces mêmes études soulignent l'écart entre le discours officiel sur l'évaluation formative et la pratique la concernant : les enseignants valorisent l'évaluation formative, mais considèrent que sa pratique est irréaliste dans le contexte scolaire actuel. Ces travaux pointent ainsi une non-cohérence entre les représentations que ces enseignants manifestent à l'égard de l'évaluation formative, la façon dont ils se représentent celle-ci et leurs pratiques déclarées.

Comment expliquer cet écart ? Quelques pistes apparaissent ici et là dans les rares travaux conduits dans ce domaine. Ainsi, les enseignants expriment la difficulté d'adapter leurs stratégies aux fonctions qu'ils désirent privilégier (difficulté à réaliser concrètement les fonctions de l'évaluation formative dans des situations précises). Ils semblent avoir des difficultés à mettre l'évaluation formative en pratique et en contexte, et nous pouvons même nous demander si le discours théorique sur l'évaluation formative est viable dans la pratique. Entre la présentation théorique de l'évaluation formative et la pratique, des écarts restent à combler.

L'évaluation formative est relativement récente, et donc s'il y a représentation de l'évaluation, celle-ci est forcément associée à quelque chose de sommatif ou de normatif. À cet effet, Scallon (2000) souligne que les élèves et les enseignants ont été rompus pendant de nombreuses années à la pratique de l'évaluation cumulative, où plusieurs éléments de vérification produisent des résultats à additionner en vue d'un bilan arithmétique. Il existe une crainte assez facile à saisir, selon laquelle, dans une telle pratique d'évaluation, le travail ou l'exercice annoncés comme « gratuit » (qui ne compte pas pour le bulletin) seront vraisemblablement négligés ou ignorés par les élèves. Une évaluation formative exclusivement axée sur la régulation, c'est-à-dire éloignée de toute contribution à l'évaluation sommative, subirait le même sort.

Ces différentes données, obtenues souvent par le recours à des questionnaires ou entrevues, posent le problème de l'écart entre le discours officiel sur l'évaluation et les pratiques évaluatives, et fournissent par ailleurs des pistes d'explication possibles. Mais qu'en est-il vraiment ? Peu d'études en didactique des mathématiques se sont intéressées aux pratiques réelles au quotidien en classe de l'enseignant à l'égard de l'évaluation formative. Il nous a semblé important d'approfondir cette question pour mieux comprendre la situation.

1.2 Objectifs de la recherche

La présente recherche vise, d'une part, à décrire et à mieux comprendre les pratiques d'évaluation formative d'enseignants de mathématiques du secondaire. Il s'agit ici de documenter les pratiques au quotidien de ces enseignants en classe à l'égard de l'évaluation formative en mathématiques : peut-on cerner dans ces pratiques le recours à une évaluation formative ? Quelle(s) forme(s) prend-elle ? D'autre part, la recherche vise également à cerner les principes sous-jacents qui guident les enseignants dans leurs pratiques évaluatives.

1.3 Cadre conceptuel

Le concept d'évaluation formative, vu sous l'angle notamment de ses différentes fonctions est envisagé pour comprendre les pratiques évaluatives formatives mises en place par les enseignants de mathématiques du secondaire. Ce cadre conceptuel nous permettra par la suite de repérer de manière rigoureuse des pratiques d'évaluation menées en classe dans l'intervention quotidienne de l'enseignant. Ces pratiques d'évaluation formative dont nous cherchons à rendre compte dans le cadre de cette recherche se réalisent en contexte, dans le jeu des contraintes dans lesquelles doit fonctionner l'enseignant (nombre d'élèves, groupe d'élèves particuliers avec ses caractéristiques, contraintes institutionnelles, etc.).

Plusieurs études (Anderson, 1989 ; Conseil Supérieur de l'Éducation, 1987, 1992 ; Dassa et Dumoulin, 1991 ; Dumoulin, 1991 ; Forgette-Giroux, Bercier-Larivière et Simon, 1996 ; Grégoire, 1996 ; McMorris et Boothroyd, 1993 ; Perrenoud, 1998 ; Scallon, 2000) montrent que l'évaluation relève davantage de la tradition que de la formation, qu'elle est bien ancrée dans l'habitus et qu'elle s'inscrit aussi dans une certaine histoire professionnelle.

Toute réflexion sur les fonctions de l'évaluation formative des apprentissages passe d'abord par une définition claire de ce type d'évaluation. Scallon (2000) donne cette définition fonctionnelle de l'évaluation formative :

> C'est un processus d'évaluation continue ayant pour objectif d'assurer la progression des individus engagés dans une démarche d'apprentissage ou de formation, selon deux voies possibles : soit par des modifications de la situation ou du contexte pédagogique, soit en offrant à chaque individu l'aide dont il a besoin pour progresser, et ce, dans chacun des cas, pour apporter, s'il y a lieu, des améliorations ou des correctifs appropriés. La « décision », c'est-à-dire la régulation, a pour objet soit la situation d'apprentissage, soit l'individu lui-même. (Scallon, 2000, p. 21).

Pour Scallon (2000), cette définition comporte plusieurs aspects qu'il faut mettre en relief :

- *un processus d'évaluation continue* : qui s'oppose à des événements qui surviendraient d'une manière sporadique, dans un effet de surprise, sans aucune planification ;

- *ayant pour objectif d'assurer la progression* : il fait allusion à l'esprit de la pédagogie de la réussite dans le cadre de laquelle le concept d'évaluation formative a d'abord été élaboré ; c'est une perspective qui est de l'ordre des intentions et non des faits réels ;
- *par des modifications de la situation d'apprentissage* : c'est l'un des objets de la régulation ou du suivi de l'évaluation formative ; les premiers écrits en évaluation formative mentionnent souvent la situation pédagogique comme objet de régulation ;
- *en offrant une aide à l'individu* : dans la plupart des systèmes éducatifs, au moment de l'évaluation certificative, chaque individu doit témoigner seul de ses connaissances, de ses habiletés et de ses compétences ; à ce titre, il est susceptible de se voir imputer ses succès comme ses échecs ; dans la définition donnée plus haut, l'individu est explicitement un objet de régulation ;
- *pour apporter des améliorations ou correctifs* : la régulation des apprentissages ne doit pas suivre un modèle strictement médical, ce qui lui conférerait une connotation souvent péjorative ; il peut s'agir de problèmes à corriger, bien sûr, mais il peut être aussi question d'orienter une progression vers des performances supérieures.

Grégoire (1996) met lui aussi l'accent sur la régulation en soulignant que l'évaluation formative a pour but de réguler le processus d'apprentissage. Elle vise à fournir au formateur et aux apprenants des informations utiles pour organiser la suite de la séquence d'apprentissage. L'évaluation formative peut se limiter à faire le point sur le niveau de la maîtrise de la compétence attendue en fin d'apprentissage. Son but est alors d'informer l'apprenant sur le chemin qui lui reste à parcourir pour atteindre le degré de maîtrise souhaité. Elle vise aussi à donner des informations utiles pour surmonter des difficultés d'apprentissage (Grégoire, 1996). Dans ce cas, l'évaluation formative prend un caractère diagnostic. Elle ne sert plus seulement à faire un bilan en cours d'apprentissage. Elle vise à comprendre les difficultés qui peuvent surgir à divers moments de ce processus d'apprentissage. Une évaluation diagnostique ne peut se satisfaire d'une information de surface qui ne concerne que le produit. Constater simplement qu'un élève ne parvient pas à effectuer correctement une addition comme, par exemple, « 27 + 15 » se révèle, en général, insuffisant pour pouvoir agir de manière efficace.

Morissette (2002) souligne que l'évaluation est formative lorsqu'elle permet de guider et d'optimiser les apprentissages en cours, sans souci de classer, de certifier, de sélectionner : seuls les effets de régulation comptent. Voilà pourquoi elle va plutôt renseigner les élèves sur les composantes du processus (savoir, savoir-faire, démarche, procédures, stratégies, attitudes, etc.), contrairement à l'évaluation sommative qui renseigne sur le produit obtenu.

Cette évaluation formative peut être faite de façon formelle ou informelle. Louis (1999) souligne la grande particularité de l'évaluation informelle, qui se déroule de façon naturelle, sans instrumentation particulière et permet de mener une intervention continue et rapide. Cette forme d'évaluation vise principalement à aider l'élève dans le processus même de son apprentissage et est généralement individuelle. L'enseignant qui observe l'élève en train de faire une activité et qui intervient pour rétroagir sur la progression de son travail par rapport au résultat attendu, qui lui fournit certains indices pour arriver à la solution d'un problème, utilise l'évaluation informelle. Il arrive souvent que l'enseignant se serve des informations recueillies de cette façon auprès d'un élève pour donner des indices supplémentaires à toute la classe sur la démarche de solution du problème. En ce qui concerne l'évaluation formelle, Louis (1999) note aussi que c'est la formalisation de l'évaluation par l'utilisation d'instruments de mesure (examens, observations, exercices, etc.), qui est généralement conseillée lorsqu'il s'agit de prendre des décisions importantes par rapport à un élève, à un groupe d'élèves, ou en vue de modifier le contexte d'enseignement. Par exemple, l'enseignant qui demande aux élèves de faire certains exercices prévus dans le cahier de mathématiques afin de constater s'ils ont bien compris les notions vues dans une leçon donnée utilise une évaluation formative formelle.

2. Méthodologie

Compte tenu de l'objet investigué, notre choix méthodologique s'est orienté vers le recours à différentes sources de prises de données permettant de rendre compte de ces pratiques évaluatives formatives, et des principes sous-jacents :

- *Une observation directe, systématique et non participante* de cinq enseignants en classe en vue d'une caractérisation des pratiques d'évaluation formative en mathématiques. Cette observation s'est appuyée sur une prise de notes systématique des pratiques observées (à l'aide d'un journal de bord) et le recueil de toutes les traces écrites (tests formatifs, exercices, problèmes donnés, devoirs, corrections, etc.) en lien avec l'évaluation.
- *Des entretiens individuels semi-structurés* : situations aménagées à partir de questions pour examiner les pratiques d'évaluation formative des enseignants telles que ces derniers les perçoivent et les appliquent au quotidien (intentions sous-jacentes/principes qui les guident).

Pour le recueil des données, l'observation a porté sur l'action des faits d'évaluation des enseignants en classe[2]. Les entretiens semi-structurés individuels venaient compléter et enrichir cette observation (triangulation des sources de données).

Cas à l'étude

L'observation et les entretiens individuels semi-structurés ont été réalisés auprès de cinq enseignants provenant d'une même commission scolaire : deux femmes et trois hommes. Ces enseignants avaient tous une formation en enseignement des mathématiques (ils avaient suivi un baccalauréat qui les prépare à enseigner les mathématiques au secondaire). Selon le cas, dans cette formation, ils ont soit eu des cours en didactique des mathématiques, soit en évaluation des apprentissages, soit les deux. Deux d'entre eux enseignaient au second cycle, les trois autres au premier cycle. Deux d'entre eux n'enseignaient que les mathématiques et les trois autres enseignaient deux matières.

Les exigences associées à une présence prolongée sur le terrain, caractéristiques de l'étude de cas, et la complexité des pratiques observées, ont restreint notre investigation à cinq cas, c'est-à-dire, cinq enseignants du secondaire. Un certain nombre de critères que nous reprendrons maintenant ont toutefois guidé le choix de ces enseignants.

Parmi ceux-ci figure le nombre d'années d'expérience (de 7 à 14 ans d'expérience au moins). En sept ans au moins de carrière, ces enseignants ont en effet vu défiler plus de trois réformes de l'enseignement des mathématiques (avant notre recherche) avec tous les changements que cela comporte pour l'évaluation des apprentissages. Ils ont donc à cet égard un certain bagage d'expériences.

Avant de présenter les résultats issus de l'analyse des données, nous reviendrons sur le cadre conceptuel ayant guidé l'analyse.

3. Cadre conceptuel de l'analyse des pratiques évaluatives

Le cadre conceptuel ayant guidé l'analyse

Une des difficultés auxquels nous avons été confronté dès le départ lors de l'observation était celle de repérer une pratique évaluative formative. En effet, de par la définition même de l'évaluation formative[3], celle-ci est fortement intégrée à l'enseignement, par la régulation qu'elle apporte dans le processus d'apprentissage. Grégoire (1996), Scallon (2000) et Gagneux (2002) s'entendent pour noter que l'évaluation formative vise à fournir à l'enseignant et à l'élève des informations utiles pour organiser la suite de la séquence d'apprentissage. Elle fait partie intégrante de l'enseignement.

Dès lors, il était important que nous nous donnions des points de repère permettant de dire que ce dont nous parlions (à partir des pratiques observées) relevait bien d'une pratique évaluative. Les balises ayant servi à délimiter, à partir des traces de l'observation et de l'entretien, une pratique d'évaluation formative proviennent du modèle de Bélair (1995).

Pour qu'on puisse affirmer qu'il s'agit d'une pratique d'évaluation formative (et non d'une pratique pédagogique), les faits d'évaluation formative dont nous rendons compte doivent rendre compte d'une *appréciation* par l'enseignant, en lien avec l'apprentissage des élèves, avec ou sans apport d'un instrument, et sans que le moment précis de l'évaluation n'ait nécessairement été décidé au préalable. Ils doivent aussi se vivre au fur et à mesure des besoins et mettre en jeu une *intention ciblée* (qui sera explicitée par l'enseignant dans l'action ou lors de l'entrevue). Ces faits d'évaluation formative peuvent partir d'une certaine observation par l'enseignant, d'une analyse d'une situation, d'une description de production d'élève, etc. et donner lieu à une *interprétation* permettant de prendre une *décision*.

Ce repérage dans le corpus de données de pratiques évaluatives s'est de plus appuyé sur deux types d'évaluation qui ont été distinguées par plusieurs auteurs (Bélair, 1995 ; Louis, 1999 ; Scallon, 2000), *l'évaluation formative formelle* et *l'évaluation formative informelle*. Nous avons donc été amené à repérer des faits d'évaluation formative dans ses aspects formels instrumentés mais également dans ses aspects informels, prenant place dans les interactions en classe entre l'enseignant et les élèves. À partir de ce cadre référence, nous avons été en mesure de délimiter rigoureusement, dans les pratiques, ce qui relève d'une pratique évaluative formative, formelle instrumentée ou informelle.

À cette étape, un codage plus fin de ces pratiques évaluatives formatives pouvait être engagé. La grille de codage (catégories émergentes issues de l'analyse) a été élaborée en nous basant sur les procédés d'analyse de contenus propres à l'analyse qualitative des données (Huberman et Miles, 1991). Cette catégorisation a été éclairée par ailleurs par certains concepts théoriques disponibles, qui ont permis de donner sens aux données. Ainsi dans cette catégorisation, nous nous sommes appuyés sur le profil d'évaluation de Bélair (1995) que nous reprenons ci-dessous, permettant de faire ressortir différentes composantes de cette pratique évaluative :

- *la fonction donnée à l'évaluation formative* (par exemple, permettre une rétroaction positive sur le progrès de l'élève, repérer des difficultés d'apprentissage) ;
- *l'intention ou les buts poursuivis* par l'enseignant dans l'évaluation formative (par exemple adapter l'enseignement aux apprentissages des élèves, apporter une remédiation, les correctifs nécessaires) ;
- *les moyens* utilisés dans l'évaluation formative (par exemple recours à des examens, à des exercices de contrôle, à des grilles d'observation, à des

échelles d'appréciation, à un journal de bord, auto-évaluation, co-évaluation) ;

- *le contenu* sur lequel porte cette évaluation (centré sur certains savoirs essentiels, l'acquisition de certaines habiletés, etc.) ;
- *les types de rétroaction* (rétroaction faisant suite à une situation dans une régulation interactive avec les élèves, rétroaction concernant chaque élève, de type diagnostic, basé sur les difficultés et les erreurs, rétroaction concernant les groupes en classe sur les progrès et la situation d'apprentissage) ;
- *les décisions prises*, d'ordre didactique ou pédagogique, au cours de l'évaluation formative (par exemple la modification des stratégies d'enseignement, corrections, remédiation à apporter).

En tant que didacticien des mathématiques, nous nous situons également dans cette analyse des pratiques d'évaluation formative dans une perspective didactique. Celle-ci nous conduira à puiser également l'analyse des données, sous chacune de ces rubriques, à certaines ressources développées dans le champs de la didactique des mathématiques (dans l'analyse des moyens et des tâches proposées aux élèves, dans la prise en compte des erreurs, etc.).

Ces différents concepts théoriques nous ont permis d'éclairer l'analyse de données en faisant ressortir, sous les catégories, une variété de pratiques évaluatives formatives qui se dégageaient dans l'action chez nos cinq enseignants.
Le travail de codage s'est donc déroulé en trois temps :

- *Repérage* à travers les données de l'observation (validées par l'entretien) de pratiques évaluatives formatives mises en place par les enseignants. Ce repérage s'est fait à partir des balises signalées précédemment (intention ciblée ; appréciation partant d'une certaine observation, analyse, description ; interprétation permettant de prendre une décision) ;
- *Identification* des pratiques évaluatives formelles et informelles, selon les balises définies par Louis (1999) ;

- *Analyse plus fine de ces pratiques* à des fins de caractérisation à partir du profil de Bélair (1995) et des ressources amenées par la didactique des mathématiques.

4. Les résultats

4.1 Les modalités d'évaluation

L'évaluation formative formelle instrumentée, explicitée comme telle par l'enseignant (L'enseignant présente celle-ci en classe aux élèves comme une évaluation formative), fait appel à des instruments tels le recours à des tests écrits, des quiz, etc.

Dans ce cas, l'analyse des résultats a permis de mettre en évidence six catégories explicitant davantage ce que recouvre cette pratique d'évaluation formative, déclarée comme telle par l'enseignant[4] :

- ***La fonction*** que donne l'enseignant à cette évaluation dans l'action, (l'intention sous-jacente poursuivie par cette évaluation formative, les buts qu'il a comme enseignant en la donnant aux élèves), par exemple, assurer la révision de la matière vue antérieurement, faire un lien avec l'évaluation sommative qu'on cherche à préparer.
- ***Les moments*** où elle apparaît (relativement à l'étape ou à l'évaluation sommative), et durant la période même où elle prend place (en début, en cours, en fin de leçon ou sur toute la période).
- ***Les moyens ou outils utilisés***, comme par exemple, des tests, des devoirs, des jeux, des quiz.
- ***Le contenu*** sur lequel porte l'évaluation formative. Par exemple on peut retrouver dans celle-ci le même contenu que celui de l'évaluation sommative, un contenu différent (problèmes plus simples, plus complexes, autres contextes). Il s'agit lorsqu'on aborde ce contenu aussi du type de tâches proposées aux élèves (par exemple exercices d'application, problèmes à résoudre, calculs, vérification de la maîtrise d'algorithmes, questions de réflexion, travail de déduction, etc.).
- ***La correction***

On retrouve sous cette rubrique plusieurs sous-catégories :

- ***Les types de correction***, par exemple, une correction collective en classe, une correction individuelle à la maison de la copie de l'élève par l'enseignant, une correction individuelle par les élèves, une correction entre élèves (correction en équipes par exemple).
- ***La responsabilité de cette correction*** : celle-ci peut incomber à l'enseignant, à l'élève ou aux élèves collectivement.
- ***Le support utilisé*** pour assurer cette correction, par exemple, des solutionnaires. Une analyse plus fine des types de solutionnaires sera alors considérée (les types de solutionnaires pouvant être différents d'un enseignant à un autre).

- ***Le suivi*** (type de rétroaction donné par l'enseignant suite à cette évaluation). On retrouve là aussi plusieurs sous-catégories :
  - ***La rétroaction donnée aux élèves*** sur cette correction, par exemple, une note, des commentaires donnés par écrit à l'élève, un retour sur les erreurs (exploitation des erreurs en classe lors de la correction).
  - ***Le suivi*** auprès des élèves par l'enseignant, par exemple, la récupération.
- ***Le suivi pour l'enseignant*** : prolongement dans l'action entreprise par la suite, par exemple, choix de nouvelles situations d'enseignement et d'apprentissage, modification de la planification par l'enseignant.

L'évaluation formative informelle basée sur la régulation des apprentissages durant l'enseignement (démarches informelles observées entre l'enseignant et les élèves dans la gestion de la situation d'apprentissage/enseignement) se subdivise, de son côté, en cinq sous-catégories qui sont :

- ***Les moyens utilisés*** dans l'action par l'enseignant, par exemple le recours à différents types de questionnements, le travail d'équipes et les échanges entre élèves qui constituent des occasions d'observations d'élèves, l'utilisation des devoirs, etc.

- ***Les moments*** où une telle régulation apparaît, par exemple au début, en cours ou en fin de leçon.
- ***Le contenu*** sur lequel porte la régulation, par exemple, des questions ouvertes posées aux élèves forçant la réflexion, des problèmes à résoudre, une solution proposée sur laquelle on demande de se prononcer (Est-elle valide ou non ? Pourquoi ?).
- ***Les types d'interactions*** favorisées, par exemple l'interaction entre l'enseignant et les élèves, entre les élèves, etc**.**
- ***La responsabilité*** de cette évaluation, qui incombe, par exemple à l'élève qui doit se prononcer sur la solution, la valider, les autres élèves ou l'enseignant.

4.2 Résultats de l'analyse interprétative des cinq cas

En regardant l'évaluation formative formelle instrumentée, de nos cinq cas étudiés, il ressort *quatre invariants* majeurs qui sont liés à la fonction de l'évaluation formative, aux moments où elle se passe, au contenu proposé aux élèves et au moyen privilégié dans cette évaluation formative formelle**.**

Quatre invariants majeurs qui se dégagent de l'analyse de l'évaluation formative formelle instrumentée

Fonction de l'évaluation formative

Tel qu'il ressort de nos observations, la fonction première de l'évaluation formative est la préparation à l'évaluation sommative. Nous parlons ici de l'évaluation formative formelle instrumentée, déclarée comme telle par l'enseignant (au moment où elle est passée). Trois indices au moins nous le montrent incontestablement. D'abord, l'ordre de passation de l'évaluation formative formelle instrumentée et de l'évaluation sommative, la première précède toujours la deuxième, et ce dans tous les cas. Elle sert de révision de la matière, permet de se situer et de se réajuster avant le sommatif, en référence à ce que nous avons observé et à ce que les enseignants disent à l'entretien. Il faut ainsi noter que l'évaluation formative formelle instrumentée sert à revenir entre autres sur les questions échouées avant l'évaluation sommative.

Les moments de l'évaluation formative

Le deuxième invariant lié au premier, est le moment de passation de l'évaluation formative. Il vient confirmer la fonction de l'évaluation formative. L'évaluation formative formelle instrumentée précède en effet toujours, et ce dans tous les cas l'évaluation sommative. Le rythme de passation est cependant différent suivant les enseignants et se passe à plusieurs moments pouvant permettre de savoir où l'élève en est avec ses apprentissages ou encore permettre de détecter les erreurs des élèves.

Le contenu de l'évaluation formative versus sommative

Le troisième invariant est le contenu de l'évaluation formative. Les contenus des évaluations formatives formelles instrumentées et des évaluations sommatives sont presque les mêmes, ils ne diffèrent que par quelques « éléments ». Il s'avère aussi que les enseignants tiennent compte des résultats de l'évaluation formative pour préparer l'évaluation sommative. Les contenus des évaluations sommatives proviennent de ceux des évaluations formatives formelles instrumentées modifiées à partir des questions échouées ou réussies par les élèves. Il arrive également que les enseignants suppriment à l'évaluation sommative les questions moins réussies à l'évaluation formative formelle instrumentée. Ce qui confirme que la finalité de l'évaluation formative est bien de préparer à l'évaluation sommative, et que l'évaluation formative est utilisée comme moyen de réguler la construction de l'évaluation sommative : il y a ici un réinvestissement et une adaptation des tâches.

Un moyen privilégié dans cette évaluation formative formelle instrumentée : le test écrit

Le quatrième et le dernier invariant est le moyen utilisé dans l'évaluation formative formelle instrumenté. Les cinq enseignants utilisent le plus souvent une évaluation de type papier-crayon. Une forme ludique est cependant aussi utilisée, à moindre importance, par un des enseignants. Le cahier de bord (ou la grille d'observation) où l'on consigne des observations sur les élèves, des remarques sur les activités réussies ou échouées, des erreurs et des stratégies des élèves est également mis en place par un des enseignants[5].

Au delà des invariants mis en évidence précédemment, des différences apparaissent également.

Des variantes dans cette évaluation formative formelle instrumentée : les types de tâches

Les types de tâches proposées aux élèves dans l'évaluation formative diffèrent d'un enseignant à l'autre et dépendent le plus souvent des contrats implicites entre l'enseignant et les élèves (par exemple, les exercices pour l'évaluation formative et les problèmes pour l'évaluation sommative, une même tâche pour les deux évaluations formative et sommative, les aspects théoriques : définitions, théorèmes, etc. pour le formatif et la résolution de problèmes pour le sommatif.)

Nous avons observé par ailleurs que les tâches des évaluations formatives formelles instrumentées sont diverses. Elles portaient sur le chapitre que les élèves venaient juste de terminer et uniquement sur celui-ci. Dans certains cas, nous avons retrouvé des tâches techniques portant sur la mémorisation et l'application de formules et de règles. Enfin, d'autres tâches étaient de nature ludique, mettant en jeu des simulations et s'articulant autour de la probabilité. D'autres enseignants mettaient l'accent sur la réflexion, le jugement critique (vrai ou faux/pourquoi ?).

Notons que les enseignants ont également un éventail de tâches qu'ils soumettent aux élèves en évaluation formative formelle instrumentée, nous pouvons citer des questions à choix multiples ; des schémas à compléter, à lire ou à interpréter ; des questions à complétion, des questions de type « vrai ou faux », etc.

Chaque enseignant a ici, ce que nous confirmera l'entrevue avec chacun d'entre eux, un cadre de référence sous-jacent (une certaine conception des mathématiques, de l'apprentissage) qui le guide dans le choix des tâches à proposer aux élèves. En effet, ils accordent aussi beaucoup d'importance à la démarche et à la justification, dès lors les tâches qu'il choisit visent à forcer ou à stimuler la réflexion. Ils intègrent les tâches ludiques dans l'évaluation formative, des simulations de manière, nous diront-ils, à concrétiser, donner sens à des concepts abstraits. Ils proposent des tâches techniques sur la mémorisation et l'application de formules et de définitions. À l'évaluation formative, certains proposent des exercices et ils réservent des problèmes pour l'évaluation sommative. Enfin, ils préfèrent pour l'évaluation formative des exercices d'application de nature algorithmique basés sur la vérification des formules et de règles apprises en classe.

Une exploration riche dans la correction qui en est faite

Nous avons remarqué que toutes les corrections de l'évaluation formative formelle instrumentée sont d'une façon générale collectives (correction collective faite par l'enseignant, en relevant dans cette correction les erreurs des élèves).

Nous pouvons signaler par ailleurs une caractéristique commune qui ressort de cette correction dans la mise en place par tous les enseignants de procédures impliquant les élèves d'une façon ou d'une autre. Diverses modalités ont ici été mises en évidence : les élèves participent à la correction (cocorrection) des copies des pairs à partir d'un solutionnaire fourni par l'enseignant ; les copies des élèves sont annotées et les élèves corrigent les erreurs importantes signalées par l'enseignant ; les élèves corrigent les copies des pairs suivant un barème proposé par l'enseignant ; échange des copies et cocorrection à partir d'une correction modèle et d'une feuille de consignes donnée par l'enseignant.

Nous avons observé que gérer une classe pendant la correction collective d'un examen, qu'il soit formatif ou sommatif, est par ailleurs très difficile : les meilleurs élèves s'ennuient et attendent uniquement les endroits où ils ont raté l'évaluation et les autres ne voient pas ce qui leur cause problème. Que peut-on dire justement de ce regard porté par les cinq enseignants sur les copies des élèves ? De leurs annotations ? Comme pour donner une rétroaction aux élèves en correction individuelle, tous les cinq enseignants rectifient eux-mêmes les erreurs faites par les élèves. De ce fait, les élèves ne participent pas à la correction de leurs erreurs. Ces enseignants déclarent clairement qu'ils n'auront pas de temps à consacrer à l'enseignement s'ils passent leur temps à corriger et à annoter les copies ou les productions d'élèves. On voit apparaître ici une contrainte importante, souvent mentionnée par les enseignants, le temps, qui constitue un frein à la mise en place de certaines pratiques évaluatives.

Un suivi à l'évaluation formative qui reste limité

Le suivi de l'évaluation formative reste limité chez nos cinq enseignants, mais il faut souligner qu'il existe lorsque les enseignants en tiennent compte dans la récupération des élèves en difficulté d'apprentissage, en retour sur les erreurs des élèves ou des questions échouées à l'évaluation formative. Ce retour sur les erreurs n'est toutefois pas toujours de même nature. Il passe d'une simple identification à un essai, dans certains cas, d'analyse/interprétation. Le suivi est aussi existant lorsque les faits et les résultats de l'évaluation formative servent à modifier le déroulement de la planification de l'enseignant à partir de l'évaluation formative et des questions des élèves.

Il faut aussi souligner que la récupération est un dispositif de suivi à l'évaluation formative. Nous avons observé que tous les élèves qui ont des difficultés de compréhension en classe ou qui ont raté leur évaluation formative formelle instrumentée sont envoyés systématiquement à la récupération avec leurs copies pour y travailler et demander de l'aide et des explications à l'enseignant. Quand l'enseignant manque de temps pour donner une quelconque explication, celle-ci est donnée à la récupération.

La lecture transversale des données met par ailleurs en évidence une présence de pratiques évaluatives formatives également dans l'informel.

Au delà de l'explicite : les pratiques informelles d'évaluation formative

Au delà de l'explicite, un potentiel d'actions est mis en évidence dans ces études de cas, qui dépasse largement ce qui précède, parmi les pratiques informelles (non nommées comme telles). Nous relevons notamment le devoir et son utilisation, le travail en équipes forçant une explicitation des démarches et une justification (qui constitue un lieu d'observation pour l'enseignant), le questionnement par l'enseignant en classe et les interactions (enseignant-élèves, élèves entre eux) permettant de se faire une idée des raisonnements des élèves, l'observation lors du travail des élèves, le recours au journal de bord, etc.

Le devoir pour compléter l'évaluation formative

Les cinq enseignants observés donnent des travaux à la maison. Ces travaux viennent après une évaluation formative formelle instrumentée non terminée en classe et souvent aussi après un module terminé. Selon ces enseignants, le devoir à la maison joue un rôle important, celui de compléter l'évaluation formative, et occupe une place dans le processus d'apprentissage en mathématiques au secondaire, et ils soulignent à l'entretien son lien avec l'évaluation formative. Voici ce que dit un des enseignants : « *D'après moi, les travaux à la maison sont nécessaires pour aider la compréhension des élèves surtout si on le donne sous forme d'exercices ou de résolution de problèmes. Ils visent à renforcer un concept nouveau vu en classe. Je les considère aussi comme une continuité de l'évaluation formative, c'est donné dans le sens formatif et ça ne compte pas. Par contre l'efficacité de ces travaux peut diminuer facilement car ces travaux peuvent devenir mécaniques et ennuyeux par la suite. Mais généralement ça marche très bien, par exemple en résolution de problèmes, il suffit que je débute le travail avec les élèves en classe pour donner une première idée aux élèves dans la façon de procéder, sinon les élèves pourront les faire et si jamais ils commettent des erreurs, ils vont les répéter et celles-ci vont s'enraciner dans la mémoire des jeunes et ils auront de la difficulté à changer leur mode de pensée. Je crois sincèrement que si les travaux ne sont pas amorcés en classe, l'efficacité de ceux-ci en prendra un coup et sera diminuée.* ».

Au départ, nous ne soupçonnions pas l'ampleur que devraient prendre les devoirs à la maison au sein du système évaluatif. À travers les déclarations des enseignants en entretien, il apparaît des liens implicites entre les travaux à la maison et l'évaluation formative, ces mêmes liens peuvent être mis en évidence lorsque les enseignants en font une correction rapide afin de faire un retour sur les erreurs des élèves, ou encore en vue d'adapter les tâches en fonction de l'avancement ou des difficultés des élèves.

Travail d'équipe et interactions

Le travail en équipe reste limité et occasionnel. Il est certes présent dans les classes observées, mais à petite dose, puisque les enseignants veulent à tout prix finir le programme et aussi parce que les classes ou les groupes-classes sont surchargés. Il y a peu d'interactions des élèves entre eux, sauf en secondaires 1 et 5 où nous avons remarqué que pendant des séquences de résolution de problèmes, les enseignants demandent aux élèves de se mettre en groupe, d'échanger les solutions et de discuter de la validité de leurs solutions respectives.

Il faut remarquer par ailleurs que les interactions enseignant-élèves sont très présentes chez les cinq enseignants. C'est à travers ce questionnement, qu'il soit collectif (ou en équipe) de l'enseignant (ou des autres élèves), qui forcera une explicitation de la démarche des élèves, que les enseignants se font une idée d'où en sont les élèves globalement.

Ces régulations (questionnement collectif, interactions entre élèves, etc.) rejoignent en partie les modes d'évaluation formative informelle proposés par Bélair (1995) et Louis (1999) :

- Observation des élèves en train de faire des activités et intervention de l'enseignant par des questions en vue de faire rectifier ou améliorer les stratégies utilisées par les élèves.
- Explicitation des élèves (l'enseignant est à la recherche des justifications), questionnement visant à susciter un conflit, ayant pour conséquence le changement de stratégies en cours de route pendant la résolution de problèmes suite à une discussion entre les élèves.
- Réaction des enseignants aux non-réponses ou aux réponses fausses des élèves en les stimulant à répondre ou à rectifier les erreurs.

- Approbation ou désapprobation de l'enseignant en demandant aux élèves la justification et l'explicitation.

À travers ce qui précède, autant dans les pratiques évaluatives formelles qu'informelles, se dessine un certain cadre de référence de l'acteur qu'est l'enseignant.

Intentions sous-jacentes de l'acteur

À partir des entretiens réalisés après l'observation, il y a lieu de mettre en évidence certains principes qui guident l'action de l'enseignant dans sa classe pendant l'évaluation formative. Ces principes viennent baliser cette action.

Principes qui viennent baliser l'action (mises en évidence par les entretiens)

En partant de la première question générale de l'entretien sur la manière dont ils voient l'évaluation formative des apprentissages, nous pouvons ressortir quelques grands principes sous-jacents à leurs pratiques.

L'évaluation formative est vue par les enseignants qui ont collaboré à la recherche comme un moyen, d'une part, permettant à l'élève de se situer au regard de ses apprentissages, de réfléchir sur ses démarches et, pour eux-mêmes d'autre part, comme leur permettant de voir comment leur enseignement a été perçu par les élèves. Ce principe, on le retrouve chez tous les enseignants, pour le premier aspect, dans le retour, la correction de l'évaluation formative formelle, à travers l'importance pour eux de faire participer les élèves dans cette correction, de les responsabiliser et, pour le second aspect, dans le suivi qu'ils donnent à cette évaluation (réajustement des tâches, de leur planification dans certains cas, identification des erreurs, etc.). Nous reviendrons sur chacun de ces aspects.

Rôle des élèves

L'évaluation formative est une occasion de faire participer les élèves au processus évaluatif par l'autocorrection et la cocorrection. Même si dans l'ensemble les enseignants observés font rarement pratiquer l'autoévaluation ou la coévaluation[7], ils reconnaissent cependant qu'elles sont des composantes de l'évaluation formative.

Place de l'évaluation formative dans l'enseignement, conception des tâches

L'évaluation formative est perçue comme faisant partie intégrante de l'enseignement. Il ressort ainsi que l'évaluation formative se fait beaucoup verbalement entre les élèves et l'enseignant à travers les questions directes ou des questions outils. Mais ceci reste une idée sous-jacente que partagent certains des enseignants. Elle reste de plus, pour tous, dans le formel, une préparation au sommatif, et l'outil privilégié est ici le test écrit.

Enfin, certains principes guident aussi la conception des tâches proposées à l'évaluation formative formelle (axer sur la réflexion, sur le jugement ; connaissance de la théorie avant de résoudre des problèmes ; exercices puis problèmes ; concrétisation et simulation par le biais de jeux.).

Des préalables en lien avec sa mise en pratique possible en classe

Les enseignants adhèrent à l'évaluation formative. Nous l'avons vu pendant nos observations et ils l'ont confirmé à l'entretien, cependant ils déplorent le manque d'intérêt et de motivation de la part des élèves à l'égard de l'évaluation formative, « *celle pour pratiquer et qui ne compte pas* ». Un des enseignant pense que les élèves de secondaire 1, étant jeunes, ne comprennent pas souvent le sens et le rôle de l'évaluation formative.

Dans le même ordre d'idées, un autre pense que les conditions d'efficacité de l'évaluation formative passent par le sens des responsabilités et de l'autonomie des élèves. Ces principes sous-jacents qui guident les enseignants sont plus en lien avec quelque chose qu'ils cherchent justement à mettre en pratique.

Certains enseignants se demandent même s'il ne faut pas accorder une note si minime soit-elle aux examens formatifs afin de motiver quelques élèves et donner du sérieux à cette évaluation tout en étant conscient que cela va à l'encontre du principe même de l'évaluation formative. D'autres enseignants pensent plutôt modifier la conception que se font les élèves de l'évaluation formative. Nous pouvons souligner ici que ces principes guident les enseignants dans la gestion de l'évaluation formative et des questions qu'ils se posent sur le fonctionnement de celle-ci en salle de classe.

Nous pouvons ressortir également le principe qui guide les enseignants dans la gestion de la correction pour impliquer les élèves et les responsabiliser dans le processus d'évaluation. Pour un enseignant, l'évaluation formative est utile car elle permet d'accroître l'autonomie et le sens des responsabilités des élèves au regard de l'évaluation formative à travers l'autocorrection et la cocorrection. Quant à un autre, l'évaluation formative est utile pour donner une rétroaction individuelle aux élèves. En ce qui concerne un dernier 'enseignant, il faut faire savoir aux élèves que s'ils échouent à l'évaluation sommative, le résultat obtenu à une évaluation formative équivalente serait pris en considération.

Difficultés signalées

Finalement, il faut remarquer que la gestion de temps et de la classe pendant l'évaluation formative en matière d'aide individuelle pose bien des problèmes aux enseignants. Ils sont conscients de l'ampleur du travail de la gestion de la rétroaction et les questions suivantes reviennent à travers les entretiens :

- Comment récupérer les élèves qui ont échoué et qui ont des difficultés d'apprentissage ?
- Où trouver le temps pour donner aux élèves une rétroaction individuelle ?
- Où trouver le temps d'enseigner la matière qui permettrait aux élèves d'atteindre les objectifs qu'il ne leur est pas possible d'atteindre, alors que le rythme imposé par l'ampleur de ces objectifs et la fréquence des évaluations ne leur permet pas ?

CONCLUSION

L'évaluation formative se trouve au cœur de l'intervention en mathématiques, si l'on reprend les orientations du programme actuel de mathématiques au secondaire et du nouveau programme (M.E.Q., 1993, 2003). L'évaluation formative se trouve être un élément central au sein du changement de paradigme en place depuis 1993, qui met l'accent sur la construction du savoir par l'élève et qui favorise le processus de résolution de problèmes.

Il s'avère que les enseignants qui ont participé à notre recherche mettent en place une diversité de pratiques évaluatives au sein de la classe. Cette évaluation formative prend place dans l'informel autant que dans le formel, et dépasse donc uniquement la préparation à l'évaluation sommative à travers le test écrit (qui reste cependant bien présente). Cette diversité dans l'informel se manifeste à travers les moyens utilisés, dans lesquels le devoir, le questionnement et les interactions avec les élèves, et entre les élèves, l'observation des élèves en activité, jouent un rôle important. Au niveau des pratiques formelles instrumentées, la diversité prend davantage place dans le type de tâches proposé, de correction qui en est faite, et le suivi : on observe donc ici un ensemble d'interventions dans l'action qui se construisent à partir de balises, de principes assurant une cohérence entre les pratiques observées et les intentions sous-jacentes. L'évaluation formative, à travers toutes ses fonctions de régulation des apprentissages, de diagnostic, d'accompagnement, de suivi est présente d'une façon générale chez les cinq enseignants observés.

Enfin, la présente étude met aussi en évidence d'autres moyens d'évaluation formative que nous n'avions pas soupçonnés, à savoir, *le devoir à domicile, le questionnement, les interactions avec les élèves, l'observation* qui ont pris une place importante dans les pratiques évaluatives formatives informelles de ces enseignants. Sous-jacents à ces pratiques, des principes d'action guident les enseignants (ce que nous montrent les entretiens) dans le choix qu'ils font.

Les pratiques d'évaluation formative informelle qui se dégagent de nos observations misent sur certains moyens repris dans le cadre de la réforme à venir (MEQ, 2003) : questionnement/interactions entre élèves et travail en équipe, observation/utilisation des jeux et utilisation du journal de bord.

Les enseignants y ont par ailleurs intériorisé une certaine conception de l'évaluation formative dont ils perçoivent bien la fonction et la finalité.

Les pratiques diverses mises en évidence ici, viables en contexte, apparaissent une source à prendre en considération si l'on veut réellement tenir compte des savoirs élaborés par les enseignants dans la mise en place de la réforme.

Notes

[1] Afin de ne pas alourdir le texte, nous conservons l'expression « enseignant » au masculin pour désigner les personnes qui oeuvrent dans le domaine de l'enseignement, sauf lorsqu'il s'agira des deux enseignantes ayant participé à notre recherche ou lorsque nous ferons référence à des citations utilisant le masculin et le féminin.

[2] Pour cela, nous le verrons ultérieurement, cela nous prenait un cadre de référence permettant de délimiter une telle pratique évaluative, et de la distinguer d'une pratique d'enseignement.

[3] Dans une pratique d'évaluation formative, des exercices de vérification conçus pour fournir un feedback peuvent prolonger des activités d'apprentissage. L'évaluation formative est intégrée à l'apprentissage (Scallon, 2004, p.25).

[4] Cette dénomination par l'enseignant ne veut pas dire pour autant que cette pratique rejoint les objectifs de ce que l'on entend habituellement par évaluation formative. Nous la traitons toutefois ici du point de vue de l'acteur qu'est l'enseignant, en explicitant ce qu'il y voit.

[5] Il faut noter que les moyens ou les outils utilisés par nos cinq enseignants sont riches et diversifiés dans le cadre des pratiques informelles d'évaluation formative, mais que le test écrit (test formatif) reste le moyen privilégié en contexte formel.

RÉFÉRENCES

Anderson, J.O. (1989). Evaluation of student achievement teacher practices and educational measurement. *Alberta Journal of Educational Research.* 35(2).

Bednarz, N. et Garnier, C. 1989. Utilisation du conflit socio-cognitif dans une pédagogie contribuant à l'élaboration des processus d'anticipation des savoirs. In Garnier (Éds). *Construction des savoirs. Obstacles et conflits*. Montréal : Éditions Agence d'Arc. 389 p.

Bélair, L. (1995). *Profil d'évaluation : une analyse pour personnaliser votre pratique.* Montréal : Éditions de la Chenelière.

Conseil Supérieur de l'Éducation. (1987). *Rapport annuel 82-83,* Sainte-Foy (Qué.) : Direction des communications du Conseil supérieur de l'éducation.

Conseil Supérieur de l'Éducation. (1992). *Évaluer les apprentissages au primaire : un équilibre à trouver*. Québec : Conseil supérieur de l'éducation.

Dassa, N. et Dumoulin, J. (1991). Les pratiques d'évaluation dans le contexte de la classe. Communication présentée lors du colloque annuel de l'ADMÉÉ. Dans *L'ADMÉÉ*. 13e session d'études. Montréal : Les Éditions de l'ADMÉÉ.

De Cotret, R. et Dassa, N. (1989). Validation d'un système informatisé de diagnostic en mathématiques au secondaire : une approche centrée sur l'analyse didactique. Dans *Mesure et Évaluation en Éducation.*

Dumoulin, J. (1991). *Étude des pratiques de l'évaluation des apprentissages réalisés par les enseignants de mathématiques et de sciences du secondaire.* Mémoire de maîtrise. Université de Montréal.

Forgettte-Giroux, R., Simon, M. et Bercier-Larivière, M. (1996). Les pratiques d'évaluation des apprentissages en salle de classe : perceptions des enseignantes et des enseignants. *Revue Canadienne de l'Éducation 21, 4*. 384-395.

Gagneux, A. (2002). *Évaluer autrement les élèves*. Paris : Presses Universitaires de France.

Grégoire, J. (1996). Quelle démarche d'évaluation diagnostique des troubles d'apprentissage en mathématiques ? Dans Grégoire, J. (éd). *Évaluer les apprentissages. Les apports de la psychologie cognitive*. Bruxelles : De Boeck Université.

Groupe Didactique. (1999). *Enseigner les mathématiques*. Poitiers : Institut de Recherche sur l'enseignement des Mathématiques.

Huberman, A.M. et Miles, M.B. (1991). *Analyse des données qualitatives.* Bruxelles : De Boeck-Université.

Louis, R. (1999). *Évaluation des apprentissages en classe : théorie et pratique*. Québec : Études vivantes Laval.

McMorris, R.F. and Boothroyd, R.A. (1993). Tests that teachers build : An analysis of classroom tests in science and mathematics. *Applied Measurement in Education, 6,* 321-342.

M.E.Q. (1993). *Programme d'études. Secondaire. Mathématiques.* Québec : Les Publications du Québec.

M.E.Q. (2003). *Politique d'évaluation des apprentissages. Version abrégée*. Québec : Les Publications du Québec.

Morissette, R. (2002). *Accompagner la construction des savoirs*. Montréal : Chenelière/McGraw-Hill.

Parent, L.; P.S. Séguin ; R. Gadbois et R. Burelle. (1993). « Croyances et attitudes des enseignants de l'ordre du primaire de la région administrative du Grand Montréal Métropolitain à l'égard de la pratique de l'évaluation formative des apprentissages de leurs élèves ». Dans *Les évaluations*. Carcassonne: AFIRSE/PUM.

Perrenoud, Ph. (1988). *L'évaluation des élèves*. Bruxelles : De Boeck Université.

Scallon, G. (2000). *L'évaluation formative*. Québec : Éditions du Renouveau Pédagogique Inc.

Van Nieuwenhoven, C. et Jonnaert, P. (1992). Une approche des représentations des enseignants du primaire à propos de l'évaluation. *Mesure et évaluation en éducation.* Vol.16. Numéros 3 et 4.

Corneille Kazadi, Ph.D.
Professeur
Université du Québec à Trois-Rivières, Québec, Canada / G9A 5H7
corneille.kazadi@uqtr.ca