ISSN No:-2456-2165

Pre-High School Students' Difficulties Related to Variables in the Lebanese Mathematics Curriculum Case Study

Manal Kiwan
Lebanese University

Nina Hayfa
Lebanese University

Abstract:- This study examined students' difficulties related to variables at grade nine using a diagnostic test designed for this purpose. It also examined the persistence of these difficulties by giving the same test to high-school students and comparing the difference in students' performance at ninth and twelfth grades.

Results revealed that students have several difficulties when dealing with variables. Some of these difficulties are not attaining the ability to determine the role of the variable, assuming that a literal symbol stands for a single value, defining variables nominally not quantitatively, not attaining the meanings of basic expressions as $2x, 3x^2\ldots$ in addition. It also revealed the persistence of these difficulties throughout high-grade levels.

Keywords:- *Variables, Literal Symbols, Difficulties, Misconceptions.*

I. INTRODUCTION

Variables are quite hard to define due to the fact that their definition and usage depends largely on the context and attempting to compact the idea of variable into one definition "oversimplifies the idea and in turn distorts the purposes of algebra" [1].

Language wise, the term *variable* indicates an element or factor that's liable to change and hence is not fixed. In the first half of the twentieth century, this indication was also applicable in Math as most textbooks at that time distinguished between constants (quantities representing single values) and variables (quantities representing many values) [2]. Ever since the late 1950s, the term "variable" was given; in mathematics; a wider range of meanings by the Mathematics Reform Movement in the United States, one of which is a constant! This contradiction between what the term represents in English and in Mathematics might confuse students.

Despite the complexity and the importance of this concept, it's rarely discussed in classes where algebra courses are given [3] and researchers reported several difficulties that students have with the uses of variables [4,5]. Some of these difficulties are: 1) defining variables nominally, not quantitatively [6], 2) assuming that variables are merely used as labels (for example supposing that m always stands for the number of meters) [7], 3) thinking of variables as representing a single or unique number [8], 4) viewing variables as representatives of positive integers only [9], 5) faultily specifying the role of variable (for example when asked to simplify the expression $3x + 5x - 24$, students create the equation $3x + 5x - 24 = 0$ and solved it to obtain $x = 3$ [2], 6) creating a link to alphabetic order when determining the value of a variable as some students as assuming that among two different literal symbol, the one that occurs first in the alphabetic order is the smaller [10] while others assign the value 1 for a, 2 for b, 3 for c and so on[11], 7) not attaining the meanings of basic expressions as $2x, 3x^2$ etc…the reason for which we see errors as $x^2 + 2x = 2x^3$ [3].

For the purpose of this study, a categorization of quantitative variables has been inspired from Phillip's [2] and Dogbey's [12] categorizations of variables as shown in the table below:

ROLE	DEFINITION	EXAMPLE
Label	Shorthand for a quantitative object (mostly in real life situations).	*A* in "*Let A be the area of triangle* ABC."
Constant	Stands for a quantity with fixed value in a specific context	π, a=3
Specific unknown	Stands for a unique unknown number that should be found.	*X* in "*Solve* $X + 5 = 8$"
Continuous unknowns	Stand for unknown values to be found yielding to interval solutions in ℝ.	*x* in "*Bound x knowing that* $1 < 2x - 4 < 7$"
Discrete unknown	Stand for unknown values to be found yielding to at least two solutions, yet the solution set is not an interval in ℝ.	*x* in "*Given that* $x^2 - 3x - 28 = 0$*,find the value of x.*"
Generalized numbers	Present patterns of numbers that give true statements	*a* and *b* in "$a + b = b + a$"
Varying quantities	Present functional relationships between quantities.	*x* and *y* in "*equation of line (d):* $y = 9x - 2$"
Parameter	A quantity that influences the output or behavior of a mathematical object but is viewed as being held constant.	*a* and *b* in "*equation of line (d):* $y = ax + b$"
Abstract symbol	Literal symbols without number referent.	*x* in "*Simplify* $\frac{x^2-1}{x-1}$ "

Table 1

II. METHODOLOGY

A. Diagnostic Test

The data for the study came from a form of 9 question survey that has been designed and administered in English and French (depending on the student's language of education). Each of the questions contained one of the nine roles of variables specified in this study.

Error analysis is the role of the diagnostic assessment in this study. It's a commonly used method for identifying students' misunderstanding in mathematics. In this approach, students' responses are reviewed in order to identify a pattern of misunderstanding.

The data obtained gives a general sense of how well students of public schools at ninth grade level understand variables. In addition, common wrong answers are documented and analyzed in order to specify the most widespread misconceptions these students have about variables.

B. Reliability of the Test

A random sample of 15 ninth- grade students was subject to the diagnostic test designed. The results obtained were analyzed according to the Kuder-Ricahrdson approach which is a method of internal analysis. Methods of internal analysis "can detect errors due to content sampling and to differences among students in testwiseness, ability to follow instructions, scoring bias, and luck in guessing answers correctly" [13]. The analysis resulted in the coefficient $\rho_{RK20} = 0.667$ which indicates that the test is reliable.

C. Validity of the Test

To ensure the content validity of the test, three issues were taken into consideration:

- The test item appropriateness: all questions of the test are appropriate for the age, grade and curriculum level of the students subject to this study. Some of the tasks are introduced to them as early as grade seven.
- Completeness of the item sample: the test is made up of nine different questions; each question is highly related to one of the nine different uses of variables specified in this study and aims to assess students' knowledge about it.
- Question complexity:
  - Questions are direct and stated in simple words thus, language wise, it's understandable to all students.
  - Familiar mathematical vocabulary is used in all questions of the test.
  - All questions of the test don't require high cognitive levels of understanding in order to ensure that the test is not complicated for students at the targeted grade levels.

D. The Test and its Analysis

The diagnostic test consists of 9 questions. Students are supposed to be familiar with these questions because they don't require high cognitive levels of understanding. Each question addresses a certain role of literal symbols. These questions and the types of variables encountered in each question are listed below along with some explanations on the way answers are classified:

- Question 1: *In a basketball game, the basket is placed 2 meters above the ground level. A player moves at a distance* u *from base of the basket. Determine* d, *the distance between the basket and the player.*

In the first question of the test, the literal symbols u and d are *varying quantities* since, as the value of u varies, the value of d varies too. Students are supposed to apply Pythagoras' theorem to obtain $d^2 = 4 + u^2$ and then apply square root to get the expression of d in terms of u, which is $d = \sqrt{4 + u^2}$. An answer is counted as correct only if the student was able to write the last expression correctly.

- Question 2: *Solve* $2x + 1 = 0$. *How many values does the variable* x *assume?*

Question 2 of the test involves the variable x in the role of unknown; in particular, *specific unknown* because this question acquires only one solution, which the students are supposed to determine. They are also supposed to state that this equation admits a single solution. An answer is counted as correct if the student was able to specify the value of x, which is $x = -\frac{1}{2}$ and to clearly state that the equation has only one solution. Failure to answer any of the two parts involved in this question leads to categorizing the answer as wrong.

- Question 3: *If x is a real number and* $0 \leq 2x + 1 \leq 1$, *what values can x assume?*

The variable in the third question of the test plays the role of *continuous unknown* because the aim of this question is to find the values of x, which are infinite. An answer is counted as correct if a student was able to state that $\frac{-1}{2} \leq x \leq 0$ after subtracting 1 and then dividing by 2 all three terms of the given inequality.

Note that in this question, students are not asked to provide the number of solutions of the inequality because students at this grade level are not familiar with infinity. If this was asked, it would have revealed some possible limitations of students' understanding of variables.

- Question 4: *If x(2x+1) = 0, how many values can the variable x assume? Explain*

In this question, students are not clearly asked to determine the values of x. Were they able to determine that this equation has two distinct values of x without solving it, the literal symbol would have been considered as abstract symbol. Yet, this is unlikely to happen because students at this grade level are not so familiar with number of solutions of quadratic equations and tend to solve the equality in order to determine its number of solutions because solving equalities is frequently encountered in the textbooks. Thus, the variable x in this question is viewed as a *discrete unknown* because when solving the equality, two distinct values of x are obtained (finite number of values). An answer is considered to be correct if it states the two distinct values of x: $x = 0$ *or* $x = \frac{-1}{2}$ and mentions that x assumes two values. Otherwise, the question is said to either have no answer or a wrong one.

- Question 5: *If* n *is any positive real number, which is larger* $\frac{n+1}{n+2}$ *or* $\frac{n+2}{n+1}$ *? Explain.*

The variable n in this question is a *generalized number* because the relation to be established among the given patterns of numbers is always true. A correct answer involves stating that $\frac{n+2}{n+1} > \frac{n+1}{n+2}$ along with a correct explanation. A correct explanation might be of the form: "$\frac{n+2}{n+1} > 1$ & $\frac{n+1}{n+2} < 1$ *so* $\frac{n+2}{n+1} > \frac{n+1}{n+2}$"

or

"the numerator of $\frac{n+2}{n+1}$ is greater than its denominator and the numerator of $\frac{n+1}{n+2}$ is less than its denominator so $\frac{n+2}{n+1} > \frac{n+1}{n+2}$".

- Question 6: *If* $3x + 2y$ *represents the total cost of 3 kilos of apples and 2 kilos of bananas, what do x and y stand for?*

The variables x and y in the sixth question of the test are labels because both are shorthand of quantitative objects. A correct answer to this question involves specifying that "*x stands for the cost of 1 kg of apples and y stands for the cost of 1 kg of bananas*". Answers of the form: "*x for apples and y for bananas*" or "*x for kilos of apples and y for kilos of bananas*" or " *x for 1 kg of apples and y for 1 kg of bananas*" or any other answer that doesn't specify the precise quantitative objects that these symbols stand for are counted as wrong. Another correct answer involves stating that "*x stands for the cost of* $\frac{2}{3}$ *kg of bananas and y stands for the cost of* $\frac{3}{2}$ *kg of bananas".*

- Question 7: *Simplify* $2x + 5y + x$.

The seventh question of the test implements a variable idea in which the variables x and y are abstract symbols because they have no number referent. A correct answer is simply obtained by adding the coefficients of x to get $3x + 5y$. Note that this is a familiar task since grade 8.

- Question 8: *If* $a = 3$ *and* $b = 4$, *what's the value of* $2a + 5b + a$*?*

The variables found in the eighth question of the test are *constants* since the values of both a and b are specified as numbers. This question is solved by plugging the specified values of a and b into the expression to obtain $2 \times 3 + 5 \times 4 + 3 = 29$. Another way to solve this question involves simplifying the given expression to $3a + 5b$ before substituting the given values of a and b to get $3 \times 3 + 5 \times 4 = 29$.

- Question 9: *Consider the lines* $(D): y = mx + 3$ *and* $(d): y = \frac{1}{2}(2mx + 1)$. *Show that these two lines are parallel.*

This question involves three variables x, y and m. The literal symbols x and y are varying quantities because the value of y varies as x varies. Besides, m is considered a *parameter* because it influences the output y but is viewed as being held constant. Students are supposed to state that the two lines have equal slopes and this slope is equal to m. If the answer didn't specify the slope, it's counted as wrong because students might be stating what seems to be a memorized sentence.

Note that one can't evaluate students' understanding of the different uses of variables without using other mathematical knowledge. Thus, if the results proved students to have difficulties solving the test; despite simplicity of the questions proposed; variables might not be the only source of these difficulties as other factors might interfere. This will be discussed in the results of the test.

III. RESULTS OF THE STUDY

A. Results of the Test for Ninth Grade Students

The results obtained on each of the questions above are classified into three different categories: correct answers, wrong answers and no answers. An answer is counted as *correct* if the student provided a correct final answer with correct analysis to all parts of the question. Besides, the case of skipping a question is classified in the *no answer* category. Otherwise, an answer is counted as *wrong*, no matter whether the mistakes committed are related to understanding of variables or other mathematical concepts. However, the obtained wrong answers are analyzed in the following chapter.

The results of the study revealed that dealing with questions containing literal symbols is challenging as many students couldn't provide correct answers to the proposed questions. An average of 76% of the students either provided wrong answers or skipped the question, with a standard deviation of 24.2%. Table 2 shows how the answers of the 96 students to each of the 9 questions on the test are distributed.

Question	1	2	3	4	5	6	7	8	9
Correct answer	1%	26%	3.1%	24%	20.8%	4.2%	46.9%	74%	14.6%
Wrong answer	58.3%	68.8%	45.9%	56.3%	55.2%	68.8%	26%	14.6%	44.8%
No answer	40.6%[1]	5.2%	51%	19.8%	24%	27.1%	27.1%	11.5%	40.7%

Table 2:- Classification of Ninth-Grade Students Responses

Some of the common wrong answers provided may indicate the misconceptions students have regarding the concept of variables. These misconceptions are listed below along with the results of each question on the survey.

- Question 1: *In a basketball game, the basket is placed 2 meters above the ground level. A player moves at a distance* u *from base of the basket. Determine* d, *the distance between the basket and the player.*

In question 1, d and u are varying quantities and students are supposed to establish the relation between these quantities. The correct answer is $d = \sqrt{4 + u^2}$ but this question proved to be the most challenging as 99% of the student sample failed to answer it correctly out of which 40.6% skipped this question.

The results revealed that 26.7% of those who provided wrong answers gave a numerical value of *d* (out of which 73% answered that *d= 2 meters*). The misconception students seem to have in this case is treating variables as constants and assuming that variables refer to only one number.

Also, 17.8% out of the 56 students who provided wrong answers replied that *d=x-2* where *x* is the height of the player. These students are at a higher level of awareness regarding variables than those who assigned a numerical value to d, but they misinterpreted what the literal symbol d stands for. Instead of considering d to be the distance between the player and the basketball, they considered it to be the difference between the height of the basket and the height of the player.

In fact, 14.3% of those who provided wrong answers were able to state that $d^2 = u^2 + 2^2$ but some of them didn't try to find d while others were unable to specify the value of d correctly. Their answers varied between $d = \sqrt{4 + u^2} = 2u$, $d = 2 + u$, $d = \sqrt{2 + u^2}$ and $d = 4\sqrt{u}$. The misconception behind these errors is related to algebraic calculations and properties regarding square root rather than variables.

Also, 14.3% of the wrong answers obtained was that $d = 2 - u$. Again, misinterpreting what the literal symbol d stands for in this question might be the reason of this error. A percentage of 12.5 out of these 14.3% chose an arbitrary value for u and determined a numerical value of d. These students might have the misconception that a literal symbol should have a specific numerical value.

ISSN No:-2456-2165

➢ Question 2: *Solve* $2x + 1 = 0$. *How many values does the variable* x *assume?*

In Question 2, the literal symbol x plays the role of a specific unknown. Students are expected to specify that $x = -\frac{1}{2}$ and state that x assumes a single value. Only 26% of the 96 students provided correct answers to this question whereas 1% of them stated that x assumes one value without solving the given equation and 42.7% specified the numerical value of x correctly but didn't specify the number of values that x assumes. This might be either due to their belief that the number of values assumed is trivially 1 since they already solved the equation and found the solution, or due to the fact that they are not used to be asked about the number of solutions of an equation, especially that the exercises of the examined textbooks didn't ask specifically about the number of solutions when solving an equation. According to this, we assume that 68.7% of the students have the ability to solve this question correctly.

The equality $x = \frac{1}{2}$ was written by 12.5% of the students. This mistake might be related either to difficulties in solving number operations or to the misconception stated by McIntyre [9] that literal symbols assume only positive values. A second look at the tests of these students reveals that 55.5% of them do have this misconception since they only provided positive values to the variables throughout the whole test.

Out of the 96 students, 2% assigned two opposite values to x and stated that it assumes two values without providing any explanation. This might be due to their disability to differentiate between $2x$ and x^2 so they assumed that this is a second degree equation. They also seem to assume that a second degree equation always has 2 opposite solutions.

➢ Question 3: *If x is a real number and* $0 \leq 2x + 1 \leq 1$, *what values can x assume?*

The literal symbol x in Question 3 is a continuous unknown since it assumes infinite number of values. Only 3 students (3.1 % of the student sample) solved the question correctly and 51% gave no answers.

Whereas 1 student replied that x can assume any value but he/she didn't try to solve the inequality. This student understands that the variable in this case is a signifier of continuous numbers but he/she couldn't determine the appropriate boundaries of the inequality.

Note that 34% of those who provided wrong answers assumed that x has a single numerical value (60% of which answered that x=0, while 26.7% with the answer $x = -\frac{1}{2}$ and the other 13.3% replied that $x = -2$). The misconception these students might have is that documented by Kieran [8] which involves viewing variables as replacements for single numbers rather than representatives of an entire set of numbers as in this case. This error might also be related to some students' difficulty in dealing with order relations, thus variables are not necessarily the source of all the mistakes found.

In fact, 10% of the wrong answers obtained involved stating that x assumes two different numerical values out of which 50% answered that $x = 0 \; or \; x = 1$. Some of these students might have assumed that the boundaries of the inequality are the values which x assumes whereas one student explained that "$-1 \leq 2x \leq 1 \; so \; 0.5 \leq x \leq 1 \; so \; x = 0 \; 0r \; x = 1$". Despite the errors this student made regarding calculations, there is a possibility that he/she has a difficulty understanding the nature of real numbers.

Reading "$x = 0 \; or \; x = 1$" makes us wonder whether students who provided this answer have the narrow view of variables as replacements of integers. For this purpose, these students' tests were reviewed. None of them provided solely integer values to variables throughout the whole test. Thus, their view to variables is not restricted to integers.

Also, 13.6% of those who provided wrong answers seemed to split the inequality into 2 inequalities by reading the + as an "*and*". They solved as if $0 \leq 2x \; and \; 1 \leq 1 \; so \; x \geq 0$ while 4.5% solved only 1 side of the inequality as follows: $2x + 1 \leq 1 \; so \; 2x \leq 0 \; so \; x \leq 0$.

➢ Question 4: *If x(2x+1) = 0, how many values can the variable x assume? Explain.*

The role of variable as discrete unknown is applied in Question 4 of the survey. Out of 96 students, 19.8% gave no answer, 24% answered this question correctly whereas another 24% only solved the equality but didn't mention that x assumes two values. thus, we assume that 48% of the students have the ability to solve this question correctly due to the same reasoning provided for the results of question 2.

In fact, 10.4% of the answers obtained to this question assigned a single value to x. The most frequent values were x=0 and x= -0.5, each occurring at a percentage of 4.2% of the answers.

➢ Question 5: *If* n *is any positive real number, which is larger* $\frac{n+1}{n+2}$ *or* $\frac{n+2}{n+1}$*? Explain.*

Question 5 involves the role of variable as generalized number. In this task, the comparison of the fraction to "1" is enough to give the answer; that means the comparison of the numerator and denominator. This procedure is taught at grades 4, 5 and 6. Besides, it's simple to compare $n + 1$ and $n + 2$ for any value of n.

A majority of 79.2% of the students failed to answer this question correctly out of which 24% provided no answer. In fact, 11.5% wrote down the right answer but gave no explanation for their choice whereas another 19.8% chose the correct answer after plugging a numerical value to n and comparing the two values of the fractions

ISSN No:-2456-2165

obtained. A misconception which might lead students to take this approach is the belief that a variable only refers to one number. Another misconception, related to logic rather variables, is that students seem to generalize what's true for a specific value of n to all values of n. Also, 4.2% answered that the first fraction is larger without giving any justification of their answer; they might have made a chaotic choice. Whereas 2.1% stated that the two fractions are equal. The misconception leading to this answer has nothing to do with variables since the reasoning of this answer was that "reciprocal numbers are equal". Only one of the explanations obtained is related to variables as one student wrote that the fractions are equal since "n doesn't end". This student seems to be aware of the limit concept as $\lim_{x\to\infty}\frac{n+1}{n+2}=\lim_{x\to\infty}\frac{n+2}{n+1}=1$ despite that both infinity and limits are not introduced at ninth grade level.

- Question 6: *If* $3x+2y$ *represents the total cost of 3 kilos of apples and 2 kilos of bananas, what do* x *and* y *stand for?*

In this question, x and y are labels. The literal identification of these symbols should be "x stands for the price of 1kg of apples and y stands for the price of 1kg of bananas". Only 4.2% of the student sample answered this question correctly whereas 27.1% skipped it.

In fact, 16.7% out of those who gave wrong answers stated that "x stands for apples and y for bananas" and 15.2% answered that "x stands for kilos of apples and y for kilos of bananas". The misconception resulting in both of these answers might be that students believe that variables are objects and not quantities as stated by Rosnick [6]. A percentage of 19.7% of the wrong answers obtained was assigning numerical values to x and y, hence considering them as fixed quantities (4.5% wrote that x=1 and y=1, 6.1% stated that x=1kg of apples and y= 1kg of bananas whereas 9.1% considered that x=3kg of apples and y=2kg of bananas). Also, 16.7% of the 66 students who gave wrong answers to this question were aware that x and y stand for quantities but couldn't be precise in specifying what these literal symbols represent (4.5% stated that "x is number of apples and y is number of bananas", 6.1% wrote that "x is the cost of apples and y is the cost of bananas" and 6.1% considered that "x is the cost of $3kg$ of apples and y is the cost of $2kg$bananas").

- Question 7: *Simplify* $2x+5y+x$.

In this question, x and y are abstract symbols. The expected correct answer is $3x+5y$. Only 46.9% students out of the 96 students were able to answer this question correctly whereas 27.1% didn't give any answer. The most common wrong answer obtained is that this expression can't be simplified at a percentage of 32 out of the wrong answers. One of the students who gave this answer explained that one can't simplify in case of addition, this is only possible in case of multiplication.

Besides, 16% of those who gave wrong answers transformed the expression into an equality by writing $2x+5y+x=0$ and then trying to find x in terms of y or vice versa. Quarter of those who took this step continued by plugging a numerical value to one of the variables and calculated a numerical value of the second. Several students seem to have the misconception that when dealing with problems involving variables, the final answer should always be numerical and contain no literal symbols. This is known in the literature review by the need for "closure". Out of the wrong answers, 8% involved plugging the value 0 to y and simplifying the expression obtained in x. Out of the 25 wrong answers obtained, 24% ended up in wrong answers of the expression. These answers varied between $xy(2+5+1), 8xy, x(2+5y), 2x^2+5y$ and $-7xy$. Writing such absurd answers implies that these subjects have not attained the meaning of basic expressions as x, x^2 , xy etc.

- Question 8: *If* $a=3$ *and* $b=4$, *what's the value of* $2a+5b+a$?

In this question, the literal symbols a and b are constants. Students proved this type of variables to be the least challenging since 74% of the 96 students answered it correctly. This percentage is the highest among the percentages of the correct answers obtained on the nine questions of the survey. Yet, despite that this task is familiar since grade seven, ten students (11% of the student sample) skipped this question.

Also, 14.3% of the wrong answers were that $26+a$ and 7.1% were $26a$. Students with this answer might have the misconception that $a=3$ and $b=4$ only once in the context of the question or a letter can have different values at the same time when it is mentioned more than once in an expression. They seem not to know that 3 is the numerical value of a for all the literal symbols a in this question.

Besides, 14.2% of the wrong answers obtained involved dealing with the literal symbols as blanks to be filled. The answers took the form $2a+5b+a=23+54+3$ or $2a+5b+a=2.3+5.4+3$.

In fact, 7.1% of students with wrong answers seemed to have, in addition to the misconception that $a=3$ only once throughout the question, the need for closure. These students wrote that $6+20+a=0$ and found the numerical value of a.

- Question 9: *Consider the lines* $(D): y=mx+3$ *and* $(d): y=\frac{1}{2}(2mx+1)$. *Show that these two lines are parallel.*

In this question, x and y are varying quantities and m is a parameter. Students are supposed to expand the equation of (d) and state that the lines are parallel since $slope\ of\ (d) = slope\ of\ (D) = m$. Out of the student sample, 40.6% gave no answer to this question whereas only 14.6% provided a correct answer.

ISSN No:-2456-2165

In fact, 32.6% of the students who didn't answer correctly failed to expand (d) correctly and 18.6% stated what seemed to be a memorable statement: the lines are parallel since they have the same slope, without any manipulation of the form of (d) or specifying the values of the slopes.

Also, 9.3% of the wrong answers obtained stated that $slope\ of\ (d) = slope\ of\ (D) = 1$. Students with this answer might have the misconception that the slope is a numerical value that contains no literal symbols; hence they chose the slope to be the coefficient of mx and not of x.

In 2.3% of the answers that were considered wrong, students solved the equation $mx + 3 = \frac{1}{2}(2mx + 1)$ but were unable to explain the result that this equation has no solution. It would have been considered a correct answer if those stated that the lines have no point of intersection hence they are parallel.

Another 2.3% of the wrong answers involved assigning a numerical value to y (mostly $y = 0$) and solving to find x in terms of m.

B. Students' difficulties Related to Variables

Dealing with questions containing literal symbols is challenging as many ninth grade students couldn't provide correct answers to the proposed questions of the diagnostic test despite that the questions are relatively easy for the ninth grade level. The common wrong answers provided indicated the following difficulties:

- Misinterpreting what a literal symbol stands for. In fact, 99% of the student sample failed to answer the following question correctly: *In a basketball game, the basket is placed 2 meters above the ground level. A player moves at a distance u from base of the basket. Determine d, the distance between the basket and the player.* Out of which 17.8% replied that *d=x-2* where *x* is the height of the player. These students misinterpreted what the literal symbol *d* stands for. Instead of considering d to be the distance between the player and the basketball, they considered it to be the difference between the height of the basket and the height of the player. Inability to view variables as representatives of entire sets of numbers (case of continuous unknowns) since only 3.1 % of the student sample solved the following question correctly: *If x is a real number and* $0 \leq 2x + 1 \leq 1$, *what values can x assume?* Whereas 51% gave no answers and 34% of those who provided wrong answers assumed that x has a single numerical value.
- Assuming that variables always refer to only one number and treating variables as constants as 34% of those who provided wrong answers for the inequality *0≤2x+1≤1* assumed that x has a single numerical value (60% of which answered that x=0, while 26.7% with the answer $x = -\frac{1}{2}$ and the other 13.3% replied that $x = -2$). The misconception these students might have is that documented by Kieran [8] which involves viewing variables as replacements for single numbers rather than representatives of an entire set of numbers as in this case. This error might also be related to some students' difficulty in dealing with order relations, thus variables are not necessarily the source of all the mistakes found.
- Considering that variables assume only positive values since $x = \frac{1}{2}$ was written by 12.5% of the students as an answer to *Solve* $2x + 1 = 0$. *How many values does the variable x assume?*. This mistake might be related either to difficulties in solving number operations or to the misconception stated by McIntyre [14] that literal symbols assume only positive values. A second look at the tests of these students reveals that 55.5% of them do have this misconception since they only provided positive values to the variables throughout the whole test.
- Not attaining the meanings of basic expressions as $2x$ and x^2 As 25 wrong answers have been obtained on the question *Simplify* $2x + 5y + x$. Out of which 24% ended up in wrong answers whose form varied between $xy(2 + 5 + 1), 8xy, x(2 + 5y), 2x^2 + 5y$ and $-7xy$. Writing such absurd answers implies that these subjects have not attained the meaning of basic expressions as x, x^2, xy etc.
- Believing that variables are objects not quantities since when trying to answer the following question: *"If $3x + 2y$ represents the total cost of 3 kilos of apples and 2 kilos of bananas, what do x and y stand for?"* Only 4.2% of the ninth grade student sample answered this question correctly, 27.1% skipped it, whereas, 16.7% out of those who gave wrong answers stated that "*x* stands for apples and *y* for bananas" and 15.2% answered that "*x* stands for kilos of apples and *y* for kilos of bananas".
- Assuming that a literal symbol can assume different values at the same time since, for the question *If* $a = 3$ *and* $b = 4$, *what's the value of* $2a + 5b + a$? 14.3% of the wrong answers were that $26 + a$ and 7.1% were $26a$. Students with theses answers might have the misconception that $a = 3$ and $b = 4$ only once in the context of the question or a letter can have different values at the same time when it is mentioned more than once in an expression. They seem not to know that 3 is the numerical value of a for all the literal symbols a in this question.

ISSN No:-2456-2165

- Dealing with literal symbols as blanks to be filled since, for latest mentioned questioned, 14.2% of the wrong answers obtained involved dealing with the literal symbols as blanks to be filled. The answers took the form $2a + 5b + a = 23 + 54 + 3$ or $2a + 5b + a = 2.3 + 5.4 + 3$.

C. Persistence of Students' Misconceptions Related to Variables

For the purpose of this study, and to check whether the earlier specified misconceptions persist throughout high school levels, the same test, which has been performed for ninth grade students, has been given to 124 high school students Public High Schools in Lebanon distributed as follows:

Grade	10	11		12		
		(scientific section)	(humanities section	Life science	General Science	Sociology and economics
Number of students	55	23	10	11	10	15

Table 3:- Distribution of High School Students

The results obtained are as follows:

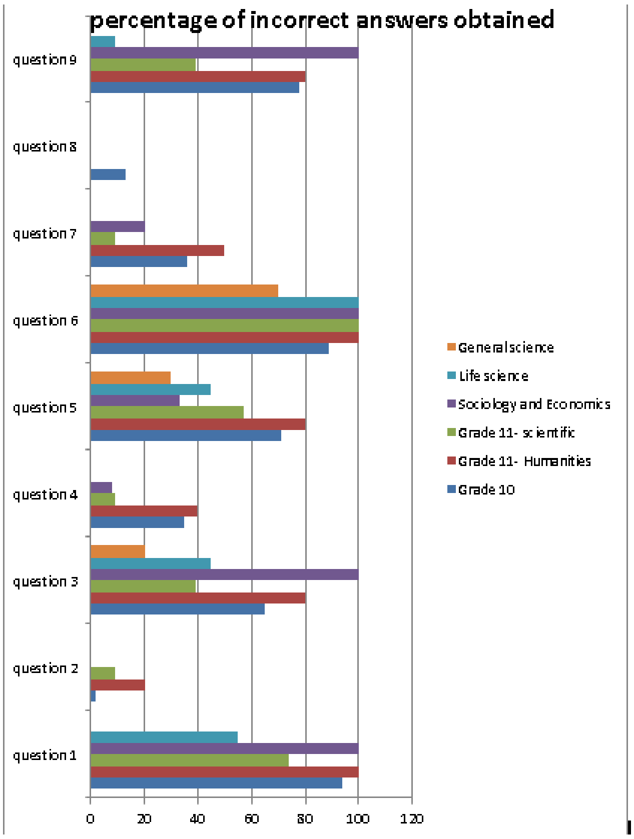

Fig 1:- percentage of incorrect answers (wrong and no answer) obtained at high school level

- *Question 1: In a basketball game, the basket is placed 2 meters above the ground level. A player moves at a distance* $\boldsymbol{u}$ *from base of the basket. Determine* $\mathbf{d}$*, the distance between the basket and the player.*

Tenth grade students seemed to have a hard time solving this question as 57% of them provided no answer and only 6% solved it correctly. Out of the wrong answers, 30% were equally distributed over the forms $d = 4 + u^2$ and $d = 2 - u$. Whereas another 30% of the wrong answers took forms: $d = 2u$, $d = \sqrt{u + 2}$, and "*d is 2 meters more to see the basket*" (at a rate of 10% each).

Besides, none of the students of the humanities section of grade 11 provided a correct answer to this question and 80% of them provided no answer at all. The two wrong answers obtained took the form $d = u$. Whereas, in the scientific section of this grade, 61% of the students stated that $d^2 = 2^2 + u^2$ but didn't continue to find d and 67% of the wrong answers mentioned that $d = 2u$.

Also, none of the students of the sociology and economics section of grade 12 provided a correct answer to this question and 20% of them didn't provide any answer at all. Out of the wrong answers obtained, 83% stated that $d = 2u$ whereas the remaining percentage was evenly distributed among the answers $d = u$ and $d = \frac{2}{u}$.

Moreover, only 36% of the students of the life science section of grade 12 provided a correct answer to this question and only one student stated that $d^2 = 4 + u^2$ without proceeding to find d. Also, 80% of the wrong answers obtained were evenly distributed among the forms $d = u + 2$ and $d = 2u$ while the remaining percentage took the form $d = 4 + u^2$.

On the other sides, students of the general sciences section proved to be well capable of solving this question as all of them answered it correctly.

- *Question 2: Solve* $\mathbf{2x + 1 = 0}$. *How many values does the variable* $\boldsymbol{x}$ *assume?*

The results obtained revealed that only one of the 55 students of grade 10 stated that $x\ assumes\ 2\ values$ without trying to solve the equation. Whereas, 20% of the students of grade 11 were unable to answer this question correctly and their answers were $x = \frac{1}{2}\ or\ x = -\frac{1}{2}$ and $x\ assumes\ 0\ values$. Also, one of the answers stated by the students of the scientific section of grade 11 was that $x\ assumes\ one\ value$ without solving the equation and another answer stated that $this\ equation\ assumes\ one\ value$: $x = -1$ whereas the remaining students were able to answer this question correctly.

On the other hand, all students of the twelfth grade with its three sections answered this question correctly.

- *Question 3:If x is a real number and 0 ≤* $\mathbf{2x + 1} \leq \mathbf{1}$*, what values can* $\boldsymbol{x}$ *assume?*

Results of the test revealed that 35% of the students of grade 10 answered this question correctly and 20% of them didn't answer it. Out of the wrong answers obtained, 24% stated that $x \leq 0$, 12% stated that $x \geq 0$ and 8% that $-1 \leq x \leq 0$.

Besides, only 20% of the students of the humanities section of grade 11 answered this question correctly and 40% of them skipped this question. Half of the wrong answers obtained stated that $x\ assumes\ no\ values$ and the rest of the wrong answers were evenly distributed among the forms $x\ assumes\ one\ value$ and $\frac{1}{2} \leq x \leq 1$.

On the other hand, 61% of the students of the scientific section of grade 11 answered this question correctly. It was noticed that 67% of the wrong answers obtained involved mentioning that $x\ assumes\ 2\ values$ whereas 11% assumed that x has only one value.

Also, 47% of the students of the sociology and economics sections failed to answer this question correctly and another 53% provided no answer to it at all. An answer of the form $0 \leq -\frac{1}{2} \leq 0$ was obtained at a rate of 29% of the wrong answers and another one stating that $0 \leq \frac{-1}{2} \leq 1$ was obtained at a percentage of 14% of the wrong answers.

Moreover, 55% of the students of the life science section of grade 12 answered this question correctly. Out of the wrong answers obtained, 20% had the form $\frac{-1}{2} \leq 2x \leq 0$ and another 40% stated that $\frac{-1}{2} \leq x \leq 0 \Rightarrow x = 0$. Also, it was stated that $x\ assumes\ 2\ values$ in 40% of the wrong answers obtained.

Students of the general sciences sections answered this question correctly at a rate of 80%. Both students who provided wrong answers stated the following: $\frac{-1}{2} \leq x \leq 0\ so\ x = 0$.

- *Question 4: If* $\boldsymbol{x(2x+1) = 0}$*, how many values can the variable x assume? Explain.*

Results revealed that 35% of the students of grade 10 couldn't give a correct a correct answer to the provided question; out of which 6% provided no answer at all. Moreover, 44% of the students who gave wrong answers assumed that x assumes only one value. Also, the answer: "$2x^2 + x = 0 \quad \Rightarrow 2x^2 = -1 \Rightarrow x^2 = \frac{-1}{2}\ (no\ solution)$" was obtained at a rate of 17% of the wrong answers.

On the other hand, 60% of the students of the humanities section of grade 11 answered this question correctly and 10% provided no answer at all. The only wrong answer obtained was that $x = 0$. Also, in the scientific section, only two students stated that $x\ assumes\ two\ values$ without providing an explanation whereas the rest of the students answered the question correctly (91% of them).

Besides, 35 out of the 36 students in the 12th grade answered the question correctly and only one student skipped it.

- *Question 5: If* $\boldsymbol{n}$ *is any positive real number, which is larger* $\frac{\boldsymbol{n+1}}{\boldsymbol{n+2}}$ *or* $\frac{\boldsymbol{n+2}}{\boldsymbol{n+1}}$*? Explain.*

Only 29% of the tenth grade students answered this question correctly as 56% of them provided wrong answers and 16% skipped answering it. It was noticed that 32% of the students who provided wrong answers got their results by substituting a specific value for n whereas 26% of them stated that $\frac{n+2}{n+1}$ is larger without explaining their choice.

Besides, only 20% of the 11th grade-students of the humanities section answered this question correctly and 40% provided no answer to it at all. Half of those who provided wrong answers assumed that $\frac{n+2}{n+1}$ is larger because "$its\ denominator\ is\ less\ than\ the\ other\ one$" whereas 25% of them stated that $\frac{n+1}{n+2}$ is larger without giving a verification for their answer.

In the scientific section of grade 11, 43% of the students answered this question correctly whereas 26% of them stated that $\frac{n+2}{n+1}$ is larger without providing an explanation. Also, 13% tried to compare the two fractions by finding their ratio but none of them could reach a final point at which the comparison is established. Moreover, 9% substituted a numerical value for n to solve this question.

On the other hand, 33% of the sociology and economics students failed to answer this question correctly and 75% of the wrong answers obtained involved substituting a numerical value for n to accomplish the comparison. Also, 45% of the students of the life science section failed to provide a logical way of comparing the two provided fractions as they randomly assigned a numerical value for n. Whereas in the general sciences section, 30% of the students tried to compare the fractions by finding their difference, but none of them could reach a final result about which is the greater.

- *Question 6: If* $\mathbf{3x + 2y}$ *represents the total cost of 3 kilos of apples and 2 kilos of bananas, what do* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *stand for?*

In the tenth grade, 89% of the students failed to answer this question out of which 32% provided no answer at all. Thirty-five percent of those who provided wrong answers stated that *"x stands for apples and y stands for bananas"* and 22% of them provided answers of the form

ISSN No:-2456-2165

"$x = 3$ *and* $y = 2$" and *"x:number of apples and y:number of bananas"* (at a rate of 11% each). Also, 22% of the wrong answers obtained included the equation *"3x+2y=5"*.

Besides, none of the students of the humanities section of grade 11 solved this question correctly as 80% of them gave wrong answers and the rest skipped it. Sixty percent of the answers obtained were equally distributed among the following forms: "*x: apples and y: bananas*", "*x: 1 kilo of apples and y:1 kilo of bananas*" and "*x=3 and y=2*".

Moreover, this question showed to be challenging for the students of the scientific section of grade 11 as well since none of them solved it correctly. The answer "*x:apples and y: bananas*" was written by 43% of the students of this grade, "*x: 1 kg of apples and y: 1 kg of bananas*" by 22% of them and "*x: mass of apples and y: mass of bananas*" by 13%.

Students of the sociology and economics section of grade 12 didn't show a better performance as all of them provided wrong answers to this question. The answer stating that *x stands for kilos of apples and y stands for kilos of bananas* was written by 47% of the students; and 20% of them considered that *x stands for 3 kilos of apples and y stands for 2 kilos of bananas.*

Besides, 55% of the students of the life science section of grade 12 provided wrong answers to this question and 67% of the answers obtained was evenly distributed over the answers "$d = 2 + u$" and "$d = 2u$".

Also, 70% of the students of the general science section of grade 12 provided wrong answers to this question and 40% of their answers stated that *"x stands for 1 kg of apples and y stands for 1 kg of bananas"*. Whereas the remaining answers had the forms "*x: kilos of apples and y: kilos of bananas*" (at a rate of 10%) and "*x=3 and y=2*" (at a rate of 20%).

- *Question 7:Simplify* $\mathbf{2x + 5y + x}$

Results revealed that 37% of the 10th grade students failed to answer this question correctly (13% skipped the question and 24% provided wrong answers). Besides, 31% of the students who gave wrong answers solved the equation "$x(2 + 5y) = 0$" and obtained that $x = 0$ *and* $y = \frac{-2}{5}$ and another 23% of them found x in terms of y after creating the equation $3x + 5y = 0$. Whereas 15% of the students who failed to answer this question correctly stated that the given expression is equal to $x(x + 1) = 5y$.

Also, half of the students of the humanities section of grade 11 answered this question correctly and 10% of them skipped it. The equation $3x + 5y = 0$ was found in half of the wrong answers obtained whereas the other half was equally distributed among the answers "$10(x^2 + y)$" and "$x(3 + 5y)$".

Moreover, results revealed that the majority of the students in the scientific section solved this question correctly (21 out of 23 students i.e. 91% of them). Both students who provided wrong answers wrote the equation $3x + 5y = 0$.

In addition, 20% of the students of the sociology and economics section of grade 12 couldn't answer this question correctly out of which 33% skipped it. All wrong answers included the equation $3x + 5y = 0$.

On the other hand, all students of the life science and general science sections answered this question correctly.

- Question 8: If $\boldsymbol{a = 3}$ and $\boldsymbol{b = 4}$, what's the value of $\boldsymbol{2a + 5b + a}$?

The results obtained revealed that 87% of the 10th grade students answered this question correctly and 9% of them skipped answering it. Half of the students who provided wrong answers wrote that $a = 6$ *and* $b = 20$ whereas the other half assumed that the given expression is equal to $6a + 9b + 3$.

On the other hand, all students of grades 11 and 12 answered this answer correctly.

- *Question 9 :Consider the lines* $(\boldsymbol{D}): \boldsymbol{y = mx + 3}$ *and*$(\boldsymbol{d}): \boldsymbol{y} = \frac{1}{2}(\boldsymbol{2mx + 1})$. *Show that these two lines are parallel.*

Results of the test revealed that 78% of the students of the 10th grade couldn't answer this question correctly as 42% of them provided wrong answers and 36% skipped it. It was noticed that 9% of the students stated that these lines are parallel since they have the same slope without determining the slopes of the lines. These answers were counted as wrong due to the fact that these students didn't show a true understanding of what they have stated. It was more like memorizing a statement than analyzing a given. Moreover, 26% of the students who provided wrong answers wrote that $y = y$ and tried to find the value of$x$.

In the humanities section of grade 11, half of the students skipped answering this question and 30% of them provided wrong answers to this question. Two third of the students whose answers were considered as wrong stated what seemed the memorized statement *the lines are parallel because they have the same slope*. Whereas the remaining third stated that "*(D)=(d)"*.

Moreover, 61% of the students of the scientific section answered this question correctly. The answer stating that *the slope of both lines is equal to 1* was written by 40% of those who provided wrong answers whereas that stating that *the slope of both lines is equal to mx* was written by 20% of them.

Also, none of the students of the sociology and economics section showed an ability to solve this question correctly as 13% of them stated the memorized statement that *the lines are parallel since they have the same slope* without determining the value of the slope and 40% skipped answering it.

Besides, only one student of the life science section stated what seemed to be a memorized statement and the others answered this question correctly.

On the other hand, all students of the general science section answered this question correctly.

The table below shows the results of the test in grade 12 with all its 3 sections included (sociology and economics, life science and general science).

Question	1	2	3	4	5	6	7	8	9
Correct answer	41.7%	100%	38.9%	97.2%	63.9%	8.3%	91.7%	100%	55.6%
Wrong answer	50%	0%	38.9%	0%	33.3%	91.7%	5.6%	0%	27.8%
No answer	8.3%	0%	22.2%	2.8%	2.8%	0%	2.8%	0%	16.7%

Table 4:- classification of the answers obtained by twelfth-grade students

The above table shows that students' performance after three academic years has become better in all of the nine questions of the test. But this improvement is not enough since especially at 12th grade level, many tasks given at ninth grade level still seem difficult for them to handle. A hundred percent of correct answers has only been obtained on questions 2 and 8. These two questions involved the use of variables as *specific unknown* and as *constant* respectively. Whereas questions involving use of variables as *varying quantities*, *discrete unknowns*, *continuous unknowns*, *abstract symbols*, *generalized numbers*, *parameters* and *labels* are still a challenge for many 12th grade students.

IV. DISCUSSION

Findings from this test are consistent with the data provided by the National Assessment of Educational Progress which indicate that many twelfth grade students are only capable of solving routine and simple algebra tasks [14].

The results obtained are compared to the results of the ninth grade students due to the fact that the test was done at the end of the academic year for ninth-grade students which implies that students start grade 10 with the documented misconceptions still persisting whereas the test was recently done to high school students at the end of February hence, the performance of tenth grade students might have developed due to teacher's interference and curriculum content or other interfering factors. Thus, comparing the results obtained from testing ninth graders to those obtained from testing 12th graders highlights students' progress in understanding the concept of variables throughout secondary levels.

It's also worth mentioning that despite the test for 12th grade students hasn't been done at the end of the academic year; their performance in tasks containing different uses of variables is unlikely to reach a satisfactory level by the end of the academic year due to the fact that in the remaining part of it, teachers are more likely to focus on the topics addressed in their textbooks and the objectives of its chapters rather than trying to deal with misconceptions that they students been holding ever since grade nine or even prior to it especially that these students are to have official exams in few months.

Hence, students' misconceptions regarding variables start as early as intermediate grades (Grades 6, 7, 8 or 9) and continue to persist throughout high school. To improve matters, more focus and effort should be done to avoid the development of students' misconceptions related to variables as early as the first grade in which literal symbols in mathematics are introduced.

REFERENCES

[1]. Usiskin, Z. (1988). Conceptions of school algebra and uses of variable. In A. F. Coxford, & A. P. Shulte (Eds.). *The ideas of algebra*, K-12 (pp. 8-19). Reston, VA: National Council of Teachers of Mathematics.
[2]. Philipp, R. A. (1992). The Many Uses of Algebraic Variables. *Mathematics Teacher*, *85*(7), 560.
[3]. Akgun, L. & Ozdemir, M.E. (2006). Students' Understanding of the Variable as General Number and Unknown: A Case Study. *The Teaching of Mathematics*. Vol. IX, 1, pp. 45–51.
[4]. Kuchemann D. (1980). The understanding of generalized arithmetic (Algebra) by secondary school children. *PhD Thesis*, University of London.
[5]. Wagner, S. (1983). What are these things called variables? *Mathematics Teacher*, 474-479.
[6]. Rosnick, P. (1982). *Student conceptions of semantically laden letters in algebra*. Amherst: University of Massachusetts, Cognitive Development Project.
[7]. Booth, L. R. (1988). Children's difficulties in beginning algebra. In A. F. Coxford & A. P.Shulte (Eds.), *The ideas of algebra*, K-12 (pp. 20-32). Reston, VA: National Council of Teachers of Mathematics.
[8]. Kieran, C. (1992). The learning and teaching of school algebra. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390-419). New York: Macmillan.
[9]. McIntyre, Z. S. (2005). An Analysis of Variable Misconceptions Before and After Various Collegiate Level Mathematics Courses. The University of Maine. *Thesis.*

[10]. Wagner, S. (1981). An Analytical framework for mathematical variables. *Proceedings of the Fifth PME Conference*, pp. 165-170. Grenoble, France.
[11]. Warren, E. (1999). The concept of a variable; gauging students' understanding. Proceedings of the XYIII PME Conference, Haifa, Palestine.
[12]. Dogbey, J. (2010). Concepts of variable in middle-grades mathematics textbooks during four eras of mathematics education in the United States. *Theses and Dissertations.* Paper 1615. http://scholarcommons.usf.edu/etd/1615.
[13]. Frisbie, D.A. (1988). Reliability of Scores from Teacher-made Tests. *Educational Measurement: Issues & Practise*, 7: 25-35. doi: 10.1111/j. 1745-3992.1988.tb 00422.X.
[14]. National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: Author.