Characterization of Mathematics tests' level of difficulty and discrimination through roughness analysis

Helmae N. Etulle[1]
Regina P. Galigao[2]
College of Education[1]
University of San Jose-Recoletos[1]
College of Arts and Sciences[2]
University of the Visayas[2]
helmae_24@yahoo.com

Submitted: August 23, 2013 Accepted: December 21, 2013

ABSTRACT

This paper is an exploratory study that deals with fractal dimensions in assessment and evaluation. Specifically, the study sought to determine how the tests characteristics: test difficulty and discrimination, may be quantified through knowledge of the fractal dimensions of the tests. As a by-product of the analysis, we may be able to identify which among the subjects in mathematics (from first year to fourth year) is found most difficult through analyzing and evaluating the ruggedness and irregularities of students' performance. The researcher made use of fractal statistics and segmentation to determine the test difficulty and the mathematical capability of the students. Results reveal that test fractal dimensions can be used as surrogate measures of both test difficulty and test discrimination indices. Both traditional test characteristics monotonically increase and decrease with increasing or decreasing fractal dimensions. High fractal dimensions indicate huge variances in student performance which are tell-tale symptoms of uneven understanding of mathematical concepts.

Keywords: *mathematical assessment, fractal dimensions, fractal analysis, test scores, mathematical performance, and mathematics education*

I. INTRODUCTION

Assessment has become a central issue in mathematics education in the new K-12 Curriculum. It has also becomes a significant field of study and learning for numerous authentic applications in real life experiences based on the outcomes-based education. Over the last decades, the importance of assessment education particularly in mathematics, had received much attention from the Department of Education, mathematics education researchers, mathematics teachers, and in many countries (Fan, 2006). The assessment standards for School Mathematics of the National Council of Teachers of Mathematics (NCTM) describes assessment as the process of gathering evidence about students' knowledge, skills, disposition towards Mathematics and of making inferences from that evidence for variety of purposes. The main purpose of mathematical education is to enable individual to apply mathematics in real-world problems (nature, society, culture). This is to help students to become mathematically literate to prepare learners for their future and occupational life. In 2005, Gallup conducted a poll that asked students to name the school subject that they considered to be the most difficult and not surprisingly, mathematics came out in the difficulty chart. Devlin (1994) states that mathematics is a science of patterns and a scientific discipline. In learning, learners must individually understand and appreciate mathematics as a science of numbers. This is an essential tool to thrive the day to day worldwide activities.

Research revealed that most students perceive mathematics as a difficult subject which has no meaning in their life (Countryman, 1999). This perception begins to develop at the primary years of schooling were students find the subject very abstract and heavily relying on concepts and methods but not able to deepen why those symbols exist. These are the realities that teachers fail to emphasize in the teaching and learning process and students fail to appreciate. This trend continues up to high school and even to college. By the time students reach to higher years, they have lost the interest in Mathematics. Worst of all, they cannot even explain some of the simple operations (Walle, 2001). The major purpose of mathematical assessment is to promote learning as a whole. The assessment is not the goal but the means to achieve those goals. The three (3) guiding principles which form the foundation of all assessment that supports effective mathematical education as cited by Desoete, et al (2004) are: (1) the content principle which means that assessment should reflect the mathematics that is most important for students to learn. (2) the learning principle which states that mathematics assessment should enhance mathematics learning and support good instructional practice. Learning mathematics must be embedded in worthwhile (engaging, educative, and authentic) problems that are part of the students' real world (Black & William, 1998). Moreover, methodologies of teaching and its approaches must enable students to reveal what they know rather than what they do not know (Coeckroft, 1992). And lastly (3) the equity principle which means that mathematics assessment should support every student's opportunity to learn the importance of mathematics. Lessons must be integrated to real experiences, explain real-life task and be practiced through authentic assessment. These principles are supported with the assessment standard in the NCTM for School Mathematics that emphasizes on the reflection of mathematical assessment to the needs of students on what to know and be able to do. Also, assessment should be learning opportunities for students to demonstrate what they know and can do mathematically. The study conducted by Steinberg, et al (2004) with research data from different levels of perspective, reveals that mathematics education, can- in general, be considered as being difficult for learner during the entire primary school career. Finally, from an educational practice point of view, the study conducted by Stock, et al (2006) points out that mathematics education is experienced as a difficult subject during a pupil's entire primary school career. Moreover, the study reveals that particular mathematics topics seem to be more difficult than others, and that some curriculum topics are experienced to be difficult in all primary and secondary learners. Furthermore, the study indicated that the choice for a specific CALP (commercially available learning packages) could matter to attain specific learning goals (Dowker, 2005). Building on the overview of the different experiences in relation to mathematics curriculum topics and the specific materials and strategies of teaching, teachers can start to develop specific interventions to circumvent the occurrence of mathematics learning difficulties or to compensate for some weaknesses in learning the mathematical concepts and skills.

From the nature of the subject up to the level of its difficulty matters a lot in the learning process. Majority of the learners have negative connotation towards the subject and often as the most difficult subject due to low grade attainment per grading. Through this study, the researcher would like to unveil the ruggedness, self-similarities, and scale in variance of test scores in mathematics to identify which among the subject in secondary level is the most difficult. This is to help learners uplift their inclination and profound interest in mathematics. This is also, for teachers to be guided on how to modify teaching strategies and its integration to real life tasks for further appreciation and higher and logical thinking.

II. CONCEPTUAL FRAMEWORK

Classical assessment and evaluation theory define "item difficulty index" and "item discrimination index" as follows: A test item difficulty index is the proportion of students obtaining the correct answer for that item to the total number of responses for that item:

1….P = R/T , where
R is the number of correct responses and T is the total number
of responses (i.e., correct + incorrect + blank responses).

Hence, the higher this index value, the lower is the difficulty, and the greater the difficulty of an item, the lower is its index. This index is counter-intuitive since it actually measures the "item easiness" rather than its difficulty. For the purposes of this study, we propose to define "item difficulty" in the following manner:

2…. $Q_j = 1 - P_j$, $j = 1,2,\ldots,T$

where T is the number of items, is now a monotone function of the difficulty level. The test difficulty, as a whole, is defined as the average of the item difficulties:

3…. Test Q = $\frac{\sum_1^T Q_j}{T}$.

Item discrimination index, on the other hand, is the item's ability to distinguish between those who know and those who do not know the answer. The traditional approach is as follows: Arrange the scores of the students in the test from highest to lowest. Obtain the upper 27% and the lowest 27% student scores. Denote by PU the unmodified difficulty index of the upper 27% for the test item and by PL the corresponding difficulty index for the lowest 27%. The item discrimination index is then:

4. $D_j = PU_j - PL_j$, $j = 1,2,\ldots T$

The test discrimination ability is defined as the average of all the item discrimination indices:

5. Test Discrimination = $\frac{\sum_1^T D_j}{T}$

The test characteristics (2) and (5) are proposed to be replaced by the concept of a fractal dimension of the test.

Fractals and Statistical Fractals. Fractals in its nature are mathematically inspired by Benoit Mandelbrot (1967; 1982), a well-known mathematician who introduced this body of knowledge to advanced mathematics. Fractals could explain myriad phenomenon and natural art around the world. It led people to understand that beyond chaos, there is order, beyond disorder, there is pattern and beyond minute thing, the whole thing could be represented. Thus, beyond irregularities of learning and human's understanding, fractals can give expounded explanations and reliable basis of such claim.

Additionally, with the method of assessment among the diverse learners, the researcher found out that their mathematical performances are not normally distributed. To analyze thoroughly the performance of the respondents, test results from different discipline of mathematics were utilized. Specifically, the data were from Mathematics 1 to Mathematics 4 of the school year 2012- 2013 of University of San Jose Recoletos. Mathematics 1 which embraced the K-12 (Spiral) Curriculum includes number concepts, number sense, measurement, algebra, and introduction to statistics. Mathematics 2 content focused on Intermediate Algebra. Mathematics 3 dealt with Geometry and lastly, Mathematics 4 was about Advanced College Algebra and Trigonometry. The test results of these subjects were constructed based on the specified competencies of the said curriculum and subject standards as mandated by the Department of Education.

Table 1. Test Scores Data.

Subjects	Number of Students	Number of Items	Highest Score	Lowest Score
MATH 1	390	100	98	5
MATH 2	385	100	100	12
MATH 3	361	90	88	11
MATH 4	353	100	96	12

Table 1 above depicts the data utilized in the study including the highest and lowest scores of each subject. Table 2, on the other hand, gives the traditional test characteristics using the formulas indicated above:

Table 2. Traditional Test Characteristics.

Subjects	No. of Students	No. of Items	Test Difficulty	Test Discrimination
Math 1	390	100	0.71	0.92
Math 2	385	100	0.65	0.87
Math 3	361	90	0.69	0.90
Math 4	353	100	0.62	0.85

III. METHODOLOGY

The test scores of the students across the various year levels were analyzed to determine if these can be appropriately modeled by fractal distributions or by the usual normal distribution. To this end, the histograms for the scores at each year level were obtained and subsequently subjected to a Kolmogorov-Smirnov test for normality.

For test scores that obey a non-normal distribution, a power law distribution or fractal distribution were fitted in accordance with the proposal of Padua et al. (2012) as follows: Mathematically, a monofractal distribution is described by the power-law probability density:

(1) $f(x;\lambda) = (\lambda-1)/\Theta\ (x/\Theta)-\lambda$, $\lambda > 1$, $x \geq \Theta$

It is shown in Padua, et.al (2013) that the maximum likelihood estimators of λ and Θ are respectively.

(2) $\lambda = 1+ n\ (\ \sum_{i=1}^{n} \log(\frac{x_1}{\theta})\ )^{-1}$

(3) $\Theta = \min \{ x1, x2,\ldots, xn \}$.

A practical approach suggested in estimating λ is to plot log f(x) versus log(x/θ) and to use the slope of the line as estimators of λ. This could be heuristically argued by taking the logarithm of both sides of (1):

(4) $\log f(x) = \log\ ((\lambda-1)/\Theta) - \lambda \log (x/\theta)$.

Moreover, the indicators of monofractality as presented by Padua, et.al (2013) fit to a fractal distribution f(x) to the quantile of the distribution G(.).

Let (x(a))be the αth quantile of G(.):

(5)… $G (x(a)) = \alpha$

At each of αth quantile of G(.), we fit a power law distribution F(.) such that:

(6)… $G (x(a)) = F(x(a)) = \alpha$,

or equivalently, obtain:

(7)...$\lambda\ (\alpha) = 1\text{- } \log⟦(1\text{-}a)⟧/(\log\ (\ x\_\alpha/\Theta))$, for all $\alpha \in (0,1)$.

Denote the empirical quantiles by $X(\alpha k)$ where $\alpha k = k/n$, $1\leq k\leq n\text{-}1$. An estimate of λ can be obtained from (7):

(8)... $\lambda = \frac{1}{n\text{-}1} \sum_{k=1}^{n\text{-}1} \lambda(ak)$.

The estimated fractal dimensions for the tests at various year levels were then correlated with the test difficulty the test discrimination indices by the usual regression analysis.

IV. RESULTS AND DISCUSSION

Figures 1-4 illustrate the histograms of the test scores for mathematics in four year levels. The histograms appear to be non-normal. This observation is formalized through a Kolmogorov-Smirnov test for normality.

Figure 1. Histogram, Math 1.

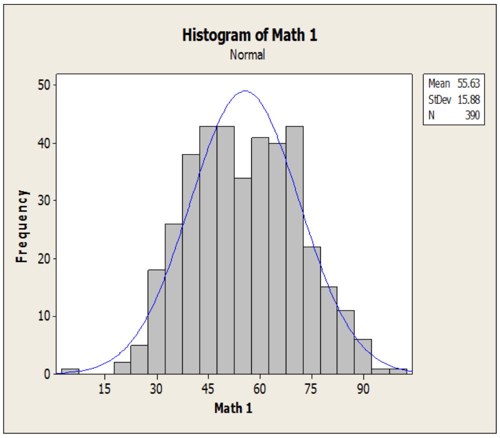


Figure 2. Histogram, Math 2.

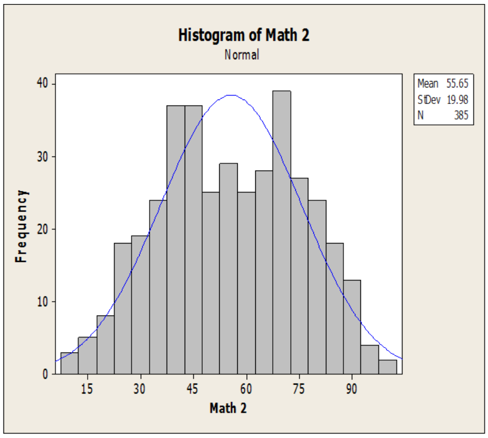


Figure 3. Histogram, Math 3.

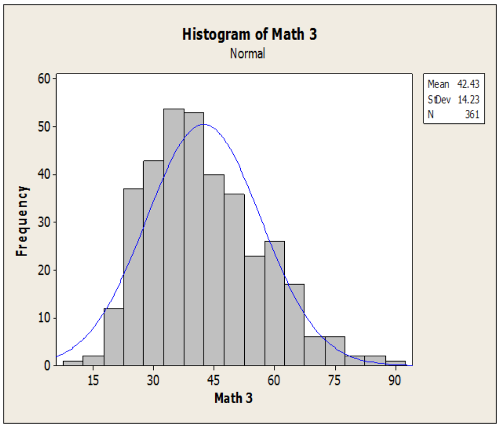
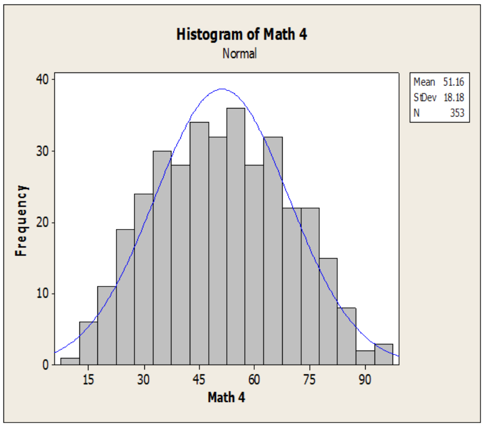


Figure 4. Histogram, Math 4.



Figures 5-8 show the formal tests for normality conducted. Significant deviations from the 45-degree lines are noted which signify deviations from normality.

Figure 5. Normality Test, Math 1.

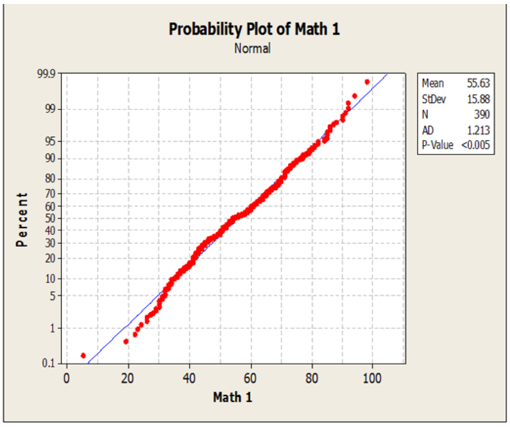


Figure 6. Normality Test, Math 2.

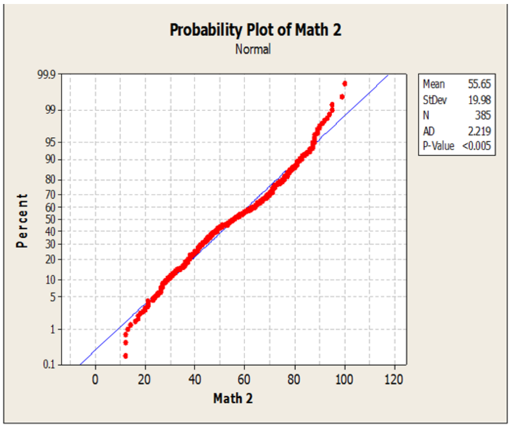


Figure 7. Normality Test, Math 3.

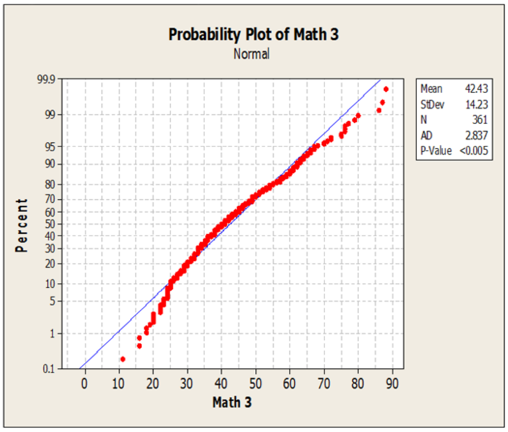


Figure 8. Normality Test, Math 4.

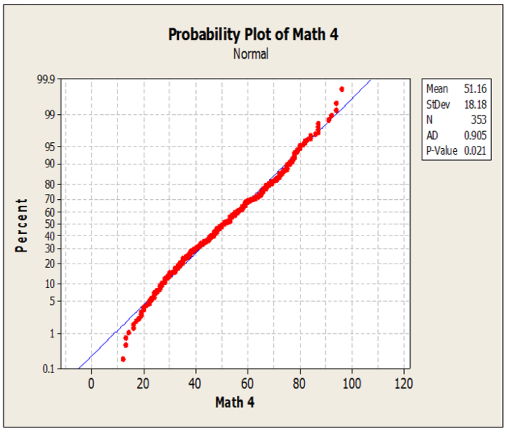
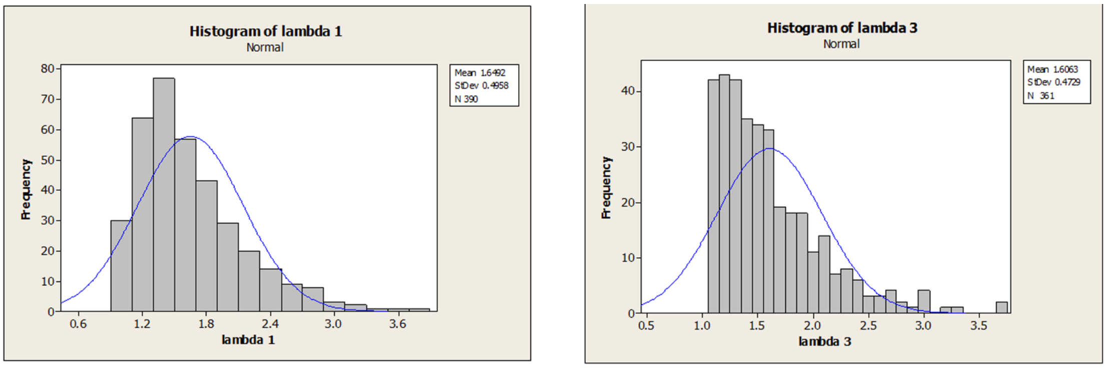


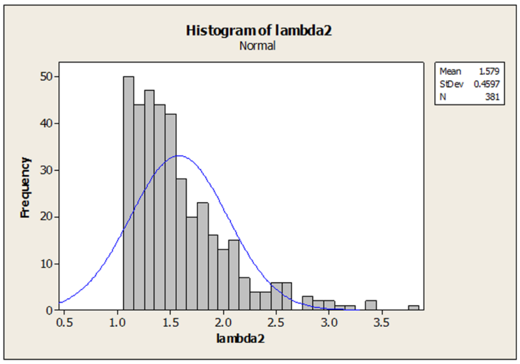
Figure 9. Histogram of the Computed Fractal Dimensions



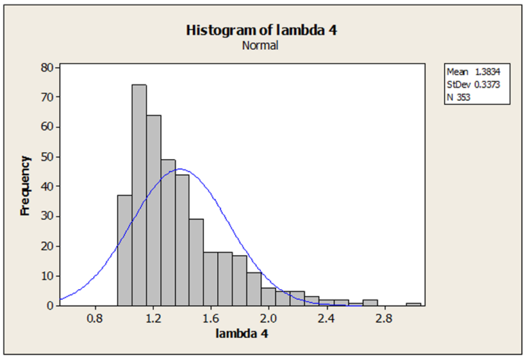
(a) (b)

Fractal Model and Analysis of Data

Fractal Dimensions. Figures 9 (a-d) show the histograms of the computed fractal dimensions for the tests. All histograms display exponential patterns consistent with the theory on fractal dimensions by Padua et al. (2012).



(c)



(d)

Table 2 below shows the descriptive statistics pertinent to the fractal dimensions of the mathematics subjects being compared.

Table 3. Descriptive Statistics of Fractal Dimensions (λ).

Variables	N	Mean	SE Mean	StDev	Minimum	Median
lambda 1	390	1.6492	0.0262	0.4958	1.0185	1.5525
lambda 2	385	1.5789	0.0235	0.4597	1.0546	1.4519
lambda 3	361	1.6063	0.0252	0.4729	1.0546	1.4889
lambda 4	353	1.3834	0.0171	0.3373	1.0034	1.2854

Table 2 shows that the subject with the highest average lambda is Math 1 while the subject with the lowest fractal dimension is Math 4.

Relating Test Characteristics and Fractal Dimensions

We proceeded to find appropriate relationships between the computed fractal dimensions of the tests and the tests characteristics of "difficulty" and "discrimination". The results are summarized in Figures 10 and 11 as well as in Tables 4 and 5.

Figure 10. Graph of Test Difficulty versus Fractal Dimension.

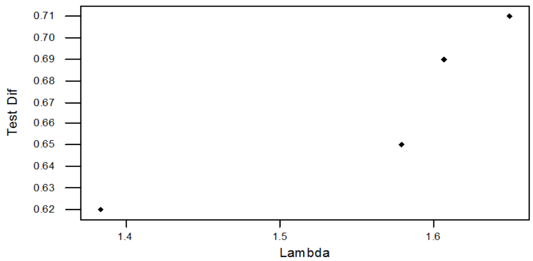


Table 4. Relationship between Test Difficulty and Fractal Dimension.

The regression equation is

Test Dif = 0.185 + 0.310 Lambda

Predictor	Coef	SE Coef	T	P
Constant	0.1849	0.1597	1.16	0.367
Lambda	0.3105	0.1025	3.03	0.094

S = 0.02089 R-Sq = 82.1% R-Sq(adj) = 73.1%

Figure 11. Graph of Test Discrimination versus Fractal Dimensions.

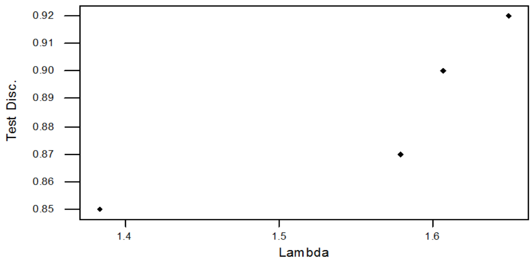


Table 5. Relationship between Fractal Dimension and Test Discrimination.

The regression equation is
Test Disc. = 0.521 + 0.234 Lambda

Predictor	Coef	SE Coef	T	P
Constant	0.5213	0.1354	3.85	0.061
Lambda	0.23395	0.08692	2.69	0.115

s = 0.01771 r-sq = 78.4% r-sq(adj) = 67.5%

The mathematics tests across the various year levels displayed high fractal dimensions. This means that students' performance in all these tests is quite erratic. The highest fractal dimension was noted in Math 1 implying that it is at this year level that huge variances in students' mathematics performance can be noted.

There appears to be a perfect matching between subjects found difficult (subjects with high difficulty indices) and subjects with high fractal dimensions. Thus, the table below shows this correspondence:

Table 6. Ranking of Fractal Dimension and Difficulty Index

Subject	Fractal Dimension	Rank	Difficulty Index	Rank
Math 1	1.6492	1	0.71	1
Math 2	1.5789	3	0.65	3
Math 3	1.6063	2	0.69	2
Math 4	1.3834	4	0.62	4

A significant linear relationship exists between fractal dimensions and difficulty indices. Empirical model obtained states: Test Dif = 0.185 + 0.310 Lambda with an R-squared value of 82%. The relationship indicates that higher fractal dimensions imply higher test difficulty.

A similar linear relationship exists between fractal dimensions and test discrimination. The empirical model states: Test Disc. = 0.521 + 0.234 Lambda. The fractal dimension of the tests explains about 72% of the variances in the test discrimination.

These results have interesting implications in measurement and evaluation. First, high variances in test performance indicate that the test instruments used have high difficulty levels. This need not necessarily imply that the subjects themselves are difficult. Second, high difficulty indices correspondingly induce high discrimination indices, both of which are summarized in terms of high fractal dimensions.

V. CONCLUSION

Test fractal dimensions can be used as surrogate measures of both test difficulty and test discrimination indices. Both traditional test characteristics monotonically increase and decrease with increasing or decreasing fractal dimensions. High fractal dimensions indicate huge variances in student performance which are tell-tale symptoms of uneven understanding of mathematical concepts.

Originality Index: 91 %
Similarity Index: 9 %
Paper ID: 3842322272
Grammarly: Checked

REFERENCES

Black, T., & William S. (1998). Literature review on classrooms, assessment and classroom learning. *Journal of Educational Research*, 29(1), 664-676.

Coeckcroft, W. H. (1992). *Mathematics counts*. London: HMSO.

Countryman, R. (1999). Individualized and meaningful instruction in arithmetic. *Journal of Educational Psychology*, 63(1), 79-84.

Desoete, A., Roeyers, H., & De Clercq, A. (2004). Children with mathematics learning disabilities in Belgium. *Journal of Learning Disabilities*, 37(1), 32-41.

Devlin, K. (1994). *The language of Mathematics. Senior Research*. CSLI.

Dowker, A. D. (2005). What works for children with Mathematical difficulties? *Research Report* no. 554. Nottingham: Department for Education and Skills.

Fan, G. T. (2006). Characteristic behavior of students with learning difficulties who have teacher-defined math weaknesses. *Journal of Learning,* 333(1), 168-179.

Gallup, A. M. (2005). *Public opinion of annual polls*. Rowman & Littlefield Publishers.

Mathematical Sciences Education Board & National Research Council (1993). *Measuring up: Prototypes for Mathematics assessment.* Washington, DC: National Academy Press.

National Council of Teachers of Mathematics (1999). *Curriculum and evaluation standards for school mathematics.* Reston, VA: NCTM.

Padua, R. N., Palompon, D. R., & Ontoy, D. S. (2012). Data roughness and Fractal Statistics. *CNU Journal of Higher Education*, 6(1), 87-101.

Steinberg, R., Haymore, J., & Marks, R. (2004, April). *Teachers' knowledge and content structuring in mathematics*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Stock, P., Desoette A., & Roeyers, H. (2006). Focusing on Mathematical disabilities: A search for definition, classification and assessment. In Randall, S. V. (Ed.) *Learning disabilities: New Research* (pp. 29-62). Hauppage, NY: Nora Science.

Walle, R. L. (2001). Diagnosis of arithmetical difficulties. *Elementary and Secondary School Journal*, 37(1), 592-600.