

# Assessment of Students' Problem-Solving Proficiency in Mathematics at Grade VI

Ghazala Noureen<sup>\*</sup>  
Itrat Irfan<sup>\*\*</sup>  
Riffatun Nisa Awan<sup>\*\*\*</sup>

## Abstract

This study was designed to develop problem solving test for class six and also assess student's problem solving proficiency. A test is constructed by using Framework of Mathematics Assessment developed by National Assessment of Educational Progress (NEAP). This framework is modified as Mathematics Problem-Solving proficiency framework by keeping in view the objectives of National Curriculum for 6<sup>th</sup> grade. Initially developed items were tried out and items for final test were selected by keeping in view the criteria of Item Response Theory (IRT). IRT provide guidance to develop the items and to construct the proficiency scales. IRT gives us a way to understand and interpret scores and also link student's achievement to their latent abilities. Final test was administered to 1500 students of four districts. Students' mean score in Algebra is greater than any other content area while performance in Ratio and Proportion is low. Therefore, students' problem-solving proficiency in the area of factors and multiples, geometry, integers, information handling, linear equations, and whole numbers is average.

**Key words:** Problem solving, mathematics proficiency, IRT

---

\* Assistant Professor, Institute of Education, Lahore College for Women  
University Lahore g\_noureen@yahoo.com

\*\* Principal Fazaia College of Education Lahore

\*\*\* Associate Professor, Department of Education, University of Sargodha.  
riffarukh@gmail.com

## **Introduction**

The most modern academic outcomes measures are made or modified from earlier measures to develop an instrument that would be more consistent, suitable, responsive and interpretable and psychometrically-sound. They become improved analytical tools afar what conventional measurement theory (or classical test theory, CTT) method can provide. One of them is item response theory (IRT). In particular, multi - dimensional IRT models are deemed suitable for modeling the complex skills demand of problem-solving tasks especially in mathematics. Problem-solving is a complete course of action rather than only a divergent issue but a process which includes understanding, formulating, reasoning, evaluating and communicating. Problem solving tasks are not routine repetitive or rote tasks (NCTM, 1989). A mathematics problem-solving task typically involves reading, identifying a solution strategy and computing (Mayer, 1992).

There are several ways to assess student's problem solving abilities. Educators and psychologists have developed different tools to assess students' problem solving skills. During recent decades assessment based on IRT model has been developed to assess student's abilities.

IRT provide guidance to develop the items and to construct the proficiency scales. IRT gives us a way to understand and interpret scores and also link student's achievement to their latent abilities. This would be helpful if the objective of an assessment is to acquaint teachers with student learning outcomes.

IRT presents student abilities and item characteristics in single mathematical scale. In this case we can develop proficiency continuum of student abilities and item difficulty on the same scale. Once student's ability is shown on the scale, we can easily make inferences about the type of task students become able to perform (Reckase, 1997).

Different intellectual processes such as reading, comprehension, computational skills and reasoning can be entrenched in a single task. Multi-component IRT models can help us to understand the sub-steps within a task, bring diagnostic process closer to cognitive processes (Embretson, 1997; Lai, 1998).

Today, IRT is being used in many large-scale testing programs. Item characteristic curve (ICC) is the essence of IRT. This curve is a description of the relationship between an examinee's ability on the trait

being measured by the item and the probability that the examinee will respond to the item correctly.

In Pakistan, Azeem (2009) developed a proficiency test of secondary level in mathematics. But there is a dire need to develop such test at elementary level. This study aimed to develop an assessment task for students to measure proficiency levels in mathematics at grade 6 by using item response theory. The main objective of this study was to measure student's mathematics ability and develop assessment task in mathematics based on IRT model and measure latent variables that can explain all the variations.

## **Literature Review**

### **Mathematics**

As per Galileo's saying the whole universe is written in mathematics language and we cannot understand it without learning its letters which are triangles, circles and other geometrical figures. Without these, one is wandering about in a dark labyrinth.

It is a conceptual system which fulfills the inner requirements of a human. It is the general study of numerals, space, shapes, quantity and the measurements. It is also called as the queen of sciences. In modern society and modern life the importance of mathematization is growing day by day. All the institutions and human activities are based and controlled numerically.

In 1950's mathematicians thought that mathematics proficiency is just to be good in computation but after 1980's they believe that actual mathematical power lies in reasoning, problem solving, linking mathematical thoughts and communicating arithmetic to others.

To develop mathematical proficiency, teachers must have an obvious visualization of the goals of teaching and what is the meaning of proficiency for the precise mathematical content which they are going to teach. It is very important to develop the understanding about the content and also know the perspective of that mathematics, how it shows the new ways to students and how it is going to boost their logical thinking.

### **Problem-Solving**

The word problem-solving, reasoning, higher order thinking, decision making are usually used simultaneously. All these terminologies refer to natural abilities and self-confidence of a student. A formal definition of problem solving is given by Anderson (1980) that "Problem

solving is defined as any goal-directed sequence of cognitive operations”. Problem-solving are those particular word problems which relate to the different branches of mathematics, also covering the aspects of reasoning and logical thinking. Hamers and Csapol (1999) wrote that ‘If we want to teach children to think then we are aiming to improve their problem-solving ability.’

### **Mathematics Proficiencies**

In early 1950’s it was believed that mathematical skills are only useful and beneficial for scientists and mathematicians. By the passage of time the need of this major subject became clearer and on a wide scale its dire need increased. Mathematical skill became essential to participate effectively in society and the financial system of 21<sup>st</sup> century.

In the last two decades, all over the world, the moment to move up learning standards and educators’ liability for success has become popular. Many countries have formulated laws to make certain that all children must get equal opportunities to come up to the high standards of mathematical proficiency. There are also mathematical standards for 1-10 in Pakistan. In schools, Mathematics has a valuable place because the aim to produce well-equipped and dependable students in a modern techno-economic society which demands sophisticated and advance mathematical skills. It also gives the opportunities of high standards jobs in the market because these job choices are formed by whether they have knowledge and capability to use mathematics (Azeem, 2013).

When concept of proficiency is taken as a whole, it was assumed that all proficiency strands are interrelated and work effectively if students are to learn effectively. Learning is a continuous process in which every mathematical proficiency strand should be harmonized with remaining strands. Elementary school teacher struggled to see whether children are making progress in all strands of mathematical proficiency.

### **The Strands of Mathematical Proficiency**

The mathematical learning’s acquire new phenomena due to changes both in society and technological world. First it was just the use of computational practice in mathematics, but on the other hand several teachers were stressing on the need for skillful performance and many of them were talking about the need for students to understand the whole process thoughtfully (Brownell, 1935). Along with the advancement basic mathematical skills such as accuracy and speed are the integral part of the mathematical learning.

In 1990's educators started giving attention to the development and improvement of 'mathematical power' which shows different mathematical goals for schools by different people at different levels. Now a day's doing well in mathematics is the requirement of talented people. However, our many circumstances have guide us to begin in a way that it becomes easy for everyone to understand and accept the objective of mathematics learning.

Now, we will discuss the kinds of intellectual changes that we should endorse in students to make them competent mathematicians. Remember that no word confine totally all traits of capability, ability, understanding and skillfulness in mathematics. But mathematical proficiency looks the right word for this which is important for all mathematical learning mathematics efficiently.

There are five strands of Mathematical proficiency:

- Conceptual understanding----- intellectual capacity of mathematic s' ideas, procedures, and associations.
- Procedural fluency----- ability to perform the action flexibly, precisely, competently, and suitably.
- Strategic competence-----potential to originate, characterize and resolve problem of mathematic.
- Adaptive reasoning-----power for rational thinking, expression, justification and validation.
- Productive disposition----- to perceive mathematics as a reasonable, helpful, and meaningful consistent tendency, enhances one's own efficiency.

The developmental process of proficiency in mathematics is based on five strands which are interwoven and mutually supporting each other.

Box 4-1

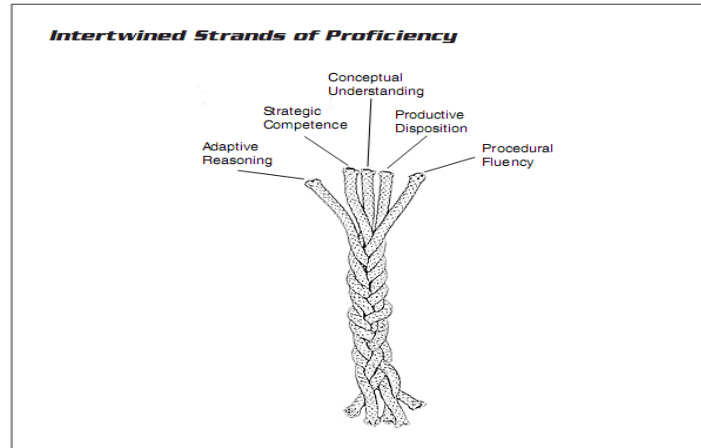


Figure 1. Proficiency Strands adopted from adding it up  
There are many proficiencies but followings are the most famous and mostly used proficiencies of mathematics.

## Methodology

Basic building block of any test is item. Mainly test construction consists of selection and development of appropriate test items. Test developers have to keep in mind certain crucial areas such as number of test items required for adequate coverage of content area, construct to be tested and reliable estimation test takers proficiencies while constructing test. One way to increase the reliability of the test is to increase the number of the items in the test. But this practice involve replication of items which actually does not fulfill the purpose of test taking because the main purpose of test taking is to get information about knowledge, skills and abilities of the test takers. So, test should consist of items having unique characteristics to measure the latent abilities of the students.

IRT plays a vital role to boost up the notion of item and test information instead of its reliability. IRT mainly deals with the test information functions which discuss the degree of exactness /level of accuracy by showing the different values of theta rather than its reliability.

An item bestowed a quality on information with the help of its piloting. The test information functions add together a large item bank with the information functions of the items on the exam.

Test information function is also helpful for the exact judging of measurement error. That's why in the reliability and validity of a test local item independence calculation is an important factor.

In Item Response Theory, RASH provides a single hidden attribute that caters all statistical reliance between test items along with differences among students. In IRT these traits are denoted by theta that differentiates items regarding their obscurity and distinguishes students according to their abilities. RM explains possibilities of students' particular answer to the question as a function of student's position on theta which describes the relationship between theta and the item.

Item response theory provides the information of a test function which explains the level of exactness at different values of theta. The test development process is cyclic in nature and divided into four major steps (Wilson, 2005).

- A     construct map
- B     selection and construction of item
- C     scoring of item
- D     measurement

### **Participants**

All public and private schools of district Lahore, Faisalabad, Sargodha and Jhung were the population of the study. Seventeen schools of above mentioned districts were selected as the sample of the study. From these seventeen schools, 1500 students were selected for test.

Item Response Theory requirement of Sample size

There are two basic requirements in order to get the combined estimation of items and ability parameter.

- A     number of items in the test
- B     number of test takers (Suen, 1990)

According to Wright and Stone (1997) minimum of 20 items and 200 test takers are required for one parameter Rasch Model and 30 and 500 test takers for two parameter Rasch Model (cited in Suen, 1997). Hambleton (1989) suggested that eighty items can be used for all models. In this study 1500 test takers were used for 30 items which sufficiently fulfill the above mentioned criteria.

### **Instrumentations**

A test is constructed by using Framework of Mathematics

Assessment developed by National Assessment of Educational Progress (NEAP). This framework is modified as Mathematics Problem-Solving proficiency framework by keeping in view the objectives of National Curriculum for 6<sup>th</sup> grade (attached as appendix).

### **Mathematical Framework**

The modified form of framework used in this study consists of following the strands.

#### Content Strands

- Integers, Algebra, Ratio and Proportion, Information Handling, Geometry.
- Procedural Knowledge:
- Procedural knowledge, conceptual understanding and problem solving.
- Mathematical Proficiency:
- Understanding of mathematical language, Mathematical logics, Mathematical association.

Initially 170 items were constructed by keeping in view the dimensions of math's proficiency framework for problem-solving strand. 50 items were aligned with National curriculum and proficiency framework. The following diagram shows the alignment of the items with National curriculum and Mathematical proficiencies.

### **Item Construction**

The aim of present study is to develop proficiency based problem solving test. Mainly, this test was developed to assess the logical, authentic and affective information about students' mathematics skillfulness. This involves decisiveness of what is to be assessed and the required items can be fabricated which can be used to test the intended knowledge, skills and perception by decreasing the effect of auxiliary ability (Gronlund & Linn, 2005). Main objective of the items is to develop premeditated answers by the learners through which we can make logical judgments (Fulcher & Davidson, 2007).

Best test item must consider the mental activity of the learner on the calculated / conscious construct which is the actual need for which this item is made. Mathematical proficiency framework deals with performance, classical content and achievement which provide guidance in the development of items.

For this purpose National curriculum was thoroughly analyzed. SLO's related to problem solving were extracted, overall 46% of the



curriculum is based on problem solving. Content wise percentages were calculated and according to these percentages number of items was selected for the pilot test.

#### Alignment to the Framework

Basically, alignment can be considered as a unit of measurement according to which different aspects of an educational system like standards, curricula and assessment act collectively to attain desirable goals ( Webb,1997b; Ananda, 2003). For the trustworthiness in education the test makers, policymakers and researchers must emphasize on the alignment of assessment and its content.

Table 1  
Description of Content Along with SLO's Percentages:

UNITS NO. AND NAMES	NO. OF SLO's	% AGES ACCORDING TO SLO's	PERCENTAGES	NO.OF ITEMS
UNIT:1 SET	$0/3 \times 100 = 0$	0%	$0/3 \times 100$	0
UNIT:2 WHOLE NUMBERS	$3/11 \times 100 = 27$	27%	$27 / 1300 \times 100$	2
UNIT:3 FACTORS AND MULTIPLES	$4/18 \times 100 = 22$	22%	$22 / 1300 \times 100$	2
UNIT:4 INTEGERS	$5/18 \times 100 = 27$	27%	$27 / 1300 \times 100$	2
UNIT:5 SIMPLIFICATION	$2/5 \times 100 = 40$	40%	$40/1300 \times 100$	3
UNIT:6 RATIO AND PROPORTION	$4/9 \times 100 = 44$	44%	$44/1300 \times 100$	4
UNIT:7 FINANCIAL ARITHMATIC	$6/8 \times 100 = 75$	75%	$75/1300 \times 100$	6
UNIT:8 ALGEBRA	$4/14 \times 100 = 29$	29%	$29/1300 \times 100$	3
UNIT:9 LINEAR EQUATION	$3/4 \times 100 = 75$	75%	$75/1300 \times 100$	6
UNIT:10 GEOMETRY	$10/10 \times 100 = 100$	100%	$100/1300 \times 100$	7
UNIT:11 PERIMETER AND AREA	$6/7 \times 100 = 86$	86%	$86/1300 \times 100$	7
UNIT:12 3D SOLIDS	$2/4 \times 100 = 50$	50%	$50/1300 \times 100$	4
UNIT:13 INFORMATION HANDLING	$1/4 \times 100 = 25$	25%	$25/1300 \times 100$	2
Total				58

### Selection of Item Formats

According to Linn and Gronlund (2005) the basic principle of test item format selection and assessment technique is to select the kind of items that provide the real expression of the intended SLO (student learning outcome).

Basic principle of item format selection is that such assessment task should be selected which provides the direct measurement of the intended learning outcomes. In proficiency testing mostly Multiple Choice Questions (MCQ's) are used internationally. By following this MCQ's format is used for the development of this test. The four options MCQ's were used. The scoring rubric was 1 for correct answer and 0 for incorrect.

### Pilot Testing

350 students were taken from six schools for pilot testing. Both rural, urban and male, female schools were selected. Test was conducted by the researcher herself to make sure the security and standard testing. ITMAN and ConQuest were used to analyze the data. Following criteria was used to select item

Difficulty Index = 0.30 to 0.80

Point bi- serial less than 0.8

Discrimination range = 0.20 to 0.60.

Table 2  
ISC (Item Selection Criteria)

Type	D index	Dis index	Analysis
1	> 0.90	any value	Description requires when teaching effected
2	0.60 --- 0.90	more than 0.20	identify extremely discerning item with reasonable intricacy. usually a good quality item
3	0.60 --- 0.90	less than 0.20	An item which cannot work. It should be handled with care or just get rid of.
4	less than < 0.60	more than 0.20	Sturdy but distinguishing. To deal with good standards, this item most likely reasonable.
5	less than 0.60	less than 0.20	Sturdy and not discriminating. This item should toss or removed.
6	Key error	Negative	An item similar to item five is a key error but answered wrongly which looks like item 1. These types of items make student crazy. A chance of verification should be given to the student. Better student should accept the teacher's key.

source: Haladyna (1997)

### Final Test

Pilot data was analyzed while using CONQUEST (IRT based software) and ITMAN (Classical theory). A selection criterion is given below.

Table 3  
Item Selection Criteria

Item Response Theory	Classical Theory
range of b is -3 --- +3	D Index 0.30 to 0.80
"Infits" means square 0.80 --- 1.20	Dis index 0.20 to 0.80
"outfit" mean square 0.80 --- 1.20	point-biserial greater 0.8

Source: Azeem, 2009.

Only those items were finalized that fulfill the above mentioned six conditions. Another issue faced by test developer is to select appropriate items for the test. How many items are required for particular test? All SLO can be considered in a single test especially in summative assessment. In content based test appropriate coverage of content is ensure the test validity. Reliability of the test and length of the test is directly proportional to each other. Guessing factor reduce the

reliability of MCQ's based test. Keeping in view above mentioned issues along with financial constraints and administrative issues 30 items were chosen for this test.

### **Preparation of Booklets**

After the finalization of test items the 1500 booklets were printed for final administration of the test. Each booklet has instruction for students and specimen example for students.

### **Test Administrator Manual**

The study was planned in such a way that test had to be administered same day in all districts. For this test administrators were required to conduct test. Due to financial constraint trained personnel of NEAS/PEAS were approached. A manual was developed to ensure the fairness and security of the test. Research supervisor and PEAS experts were consulted for manual development. This manual was finalized by incorporating the suggestions of experts.

### **Test Administration**

The test was administered on 30<sup>th</sup> May 2013. It was mandatory for all test administrators to attend morning assembly. The selected time for test administration was 9 am. The test was taken by the selected students. Administrators were asked to read instructions for students which were given in the test booklets. Students were asked to fill up first page and solved the specimen examples. The time allocated was 30 minutes for 30 items. Material was received from test administrators through courier and the expenditures were sent back through mobile credit.

## **Results**

Final data was analyzed using IRT based software CONQUEST. Final data from Microsoft Excel file was exported to SPSS file and the converted to file compatible with Conquest. Item-person maps of each strand of table of specification were generated. Weighted likelihood estimation was used for conversion of students' abilities into scaled scores using following formula.

Scaled Score on 1-1000 Scale =  $500 + 100 X (\theta)$

Where  $\theta$  = Students' ability on latent Scale

500 = Mean of Scale 1-1000

100 = SD

Item-Person maps explore the difficulty levels of items and students' ability to solve math problems on the same latent scale.

This study was designed to develop a proficiency test based on IRT. The main objective of this test development was to evaluate students' problem solving proficiency in Mathematics at level VI. It is assessed by item response theory (IRT). First a proficiency test consist of 50 items was developed for piloting. This test was based on multiple choice (MCQ) questions. Test was associated with benchmarks and standards of Mathematics for 6<sup>th</sup> class and its framework was aligned with National Assessment Examination Program (NAEP). This framework discussed how to measure the Mathematical problem-solving. The primary data which is gathered by piloting was analyzed by software's ITEMEN and CONQUEST for both CTT and IRT. After pilot test the items were selected and finalized for the final data collection. Final test was comprised of 30 MCQs questions. Multiple choice item formats was used. The final concluded proficiency test directed in different schools of Punjab province and final data was gathered from the students' of class VI. Missing data was of only 2% and it was not considered in the analysis of final data. Sampling is done by simple two stage cluster sampling design for the reason that it provides an opportunity to carry out analysis of data aggregation at higher levels. Sample of 1500 students were taken for the study of IRT based analysis. In order to understand complete design traditional statistics were used. Only minimal data was missing. Scaled used to analyze responses expected that a single answer was given for each question.

Table4  
Mean and SD for student's problem-solving proficiency

Strand	No	M	SD
Whole Numbers	1500	471.29	106.06
Integers	1500	526.55	112.91
Simplification	1500	527.74	113.22
Factors and Multiples	1500	573.67	123.82
Ratio and Proportion	1500	419.47	107.08
Algebra	1500	576.07	143.34
Area and Perimeter	1500	538.36	129.38
Geometry	1500	479.60	134.70
Volume and Surface Area	1500	493.56	109.69
Information Handling	1500	504.98	109.77
Linear Equation	1500	503.93	142.08

Mean is the mean scaled score of each student's problem-solving proficiency calibrated with an arbitrary scale 0-1000 with mean 500. Students mean score in Algebra is greater than any other content area while have lowest score in Ratio and Proportion. Following are the item response theory (IRT) based analysis of each content area of the math test generated by Conquest software.

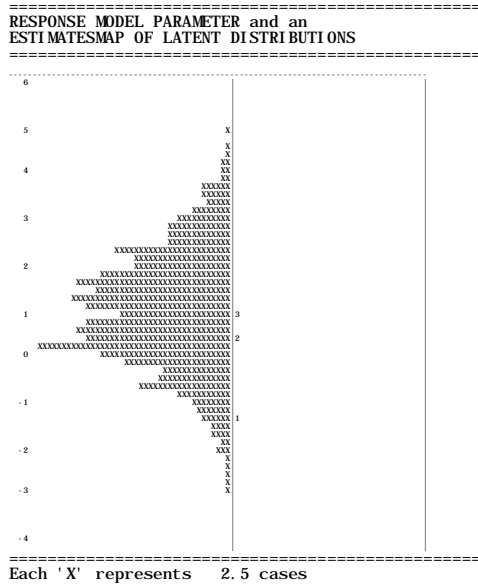
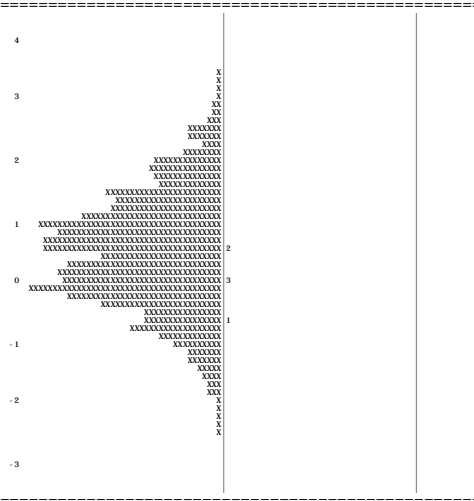


Figure 2. Algebra

Item-person map shows that students' ability to solve problem in algebra is between -1.5 to 1. Therefore students' problem-solving proficiency in the area of algebra is average.

RESPONSE MODEL PARAMETER ESTIMATES AND MAP OF LATENT DISTRIBUTIONS RESPONSE

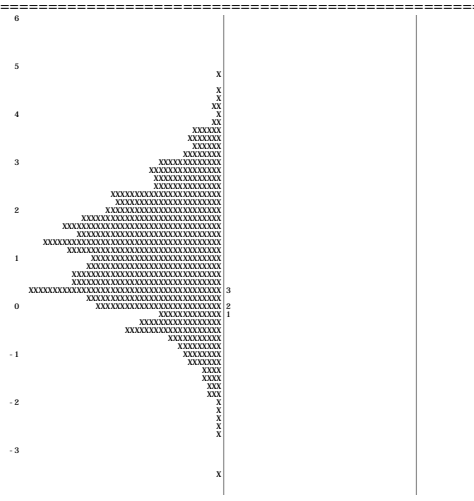


Each 'X' represents 2.3 cases

Figure 3. Area and perimeter

Item-person map shows that students' ability to solve problem in area and perimeter is between -0.5 to 0.5 Therefore students' problem-solving proficiency in area and perimeter is average.

RESPONSE MODEL PARAMETER ESTIMATES AND MAP OF LATENT DISTRIBUTIONS



Each 'X' represents 2.4 cases

Figure 4. Factors and multiples

Item-person map shows that students' ability to solve problem in factors and multiples is about 0. Therefore students' problem-solving proficiency in Factors and multiples is average.

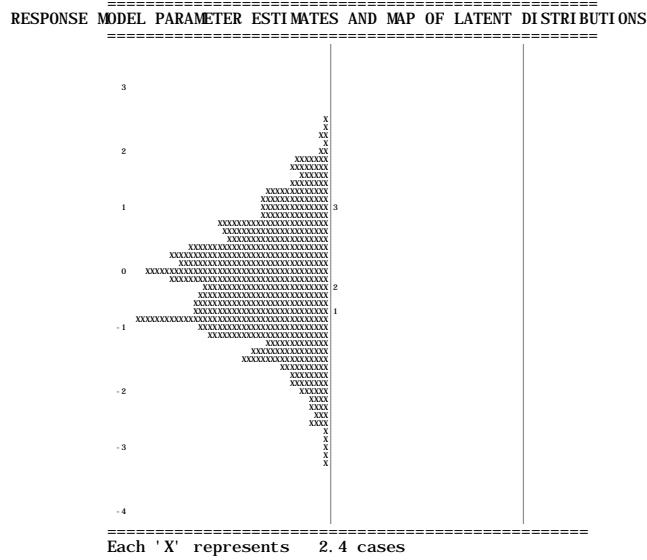


Figure 5. Geometry

Item-person map shows that students' ability to solve problem in geometry is between -1 to 1. Therefore students' problem-solving proficiency in geometry is average.

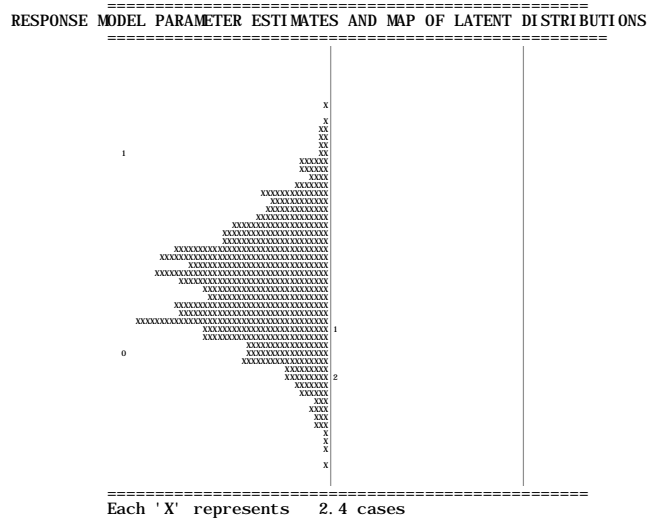


Figure 6. Integers

Item-person map shows that students' ability to solve problem in integers is about 0. Therefore students' problem-solving proficiency in integers is average.

=====

RESPONSE MODEL PARAMETER ESTIMATES AND MAP OF LATENT DISTRIBUTIONS

=====



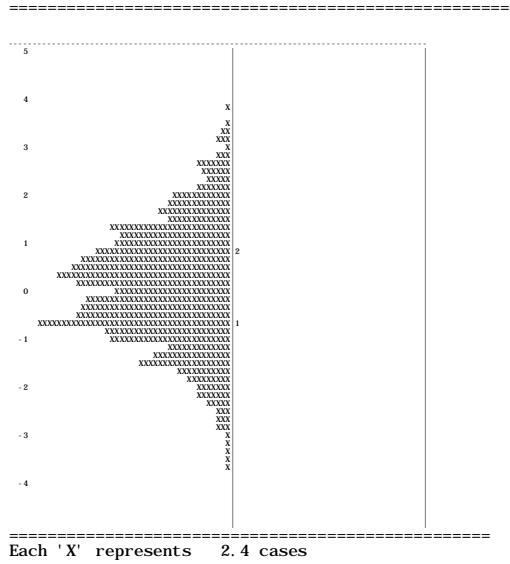


Figure 7. Linear equation

Item-person map shows that students' ability to solve problem in linear equations is about between -1 to 1. Therefore students' problem-solving proficiency in linear equations is average.

RESPONSE MODEL PARAMETER ESTIMATES AND MAP OF LATENT DISTRIBUTIONS

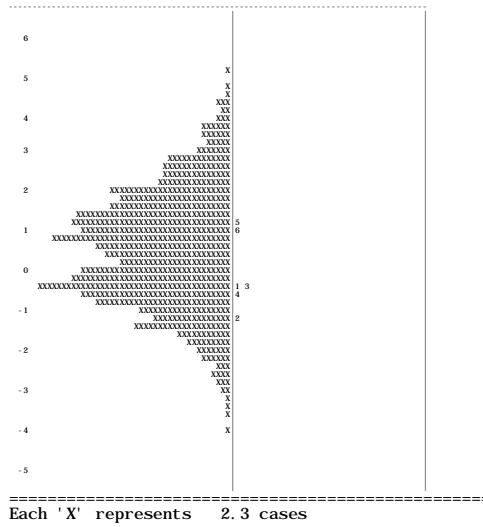
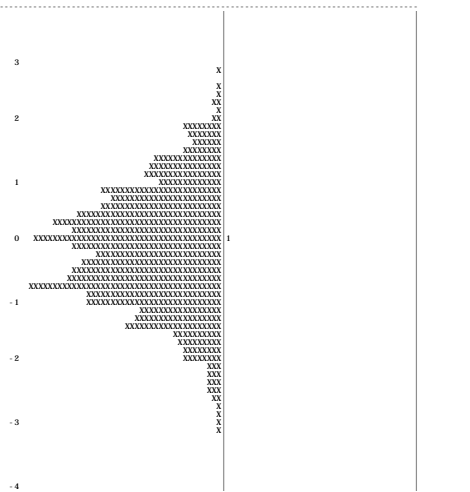


Figure 8. Simplification



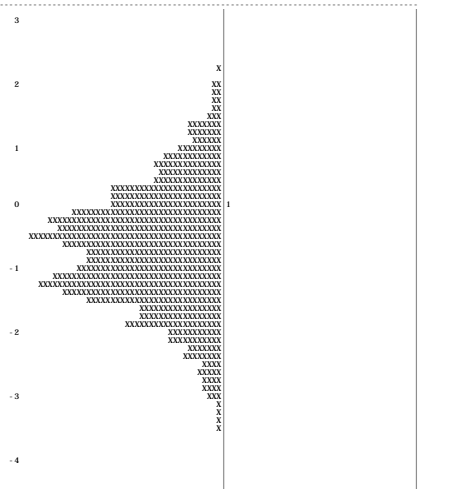
RESPONSE MODEL PARAMETER ESTIMATES AND MAP OF LATENT DISTRIBUTIONS



Each 'X' represents 2.3 cases

Figure 10. Volume and surface area Item-person map shows that students' ability to solve problem of volume and surface area is about 0. Therefore students' problem-solving proficiency in volume and surface area is average.

AND RESPONSE MODEL PARAMETER ESTIMATES AND MAP OF LATENT DISTRIBUTIONS



Each 'X' represents 2.2 cases

Figure 11. Whole numbers



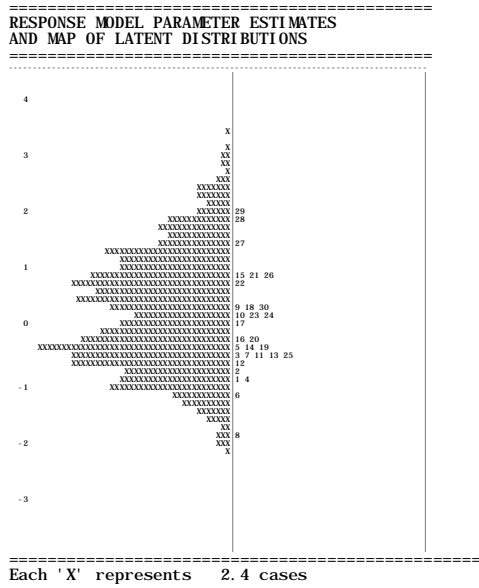


Figure 13. Overall

Item-person map shows that students' ability to solve problem in mathematics is between -2 to 2. Therefore students' overall problem-solving proficiency in mathematics is good.

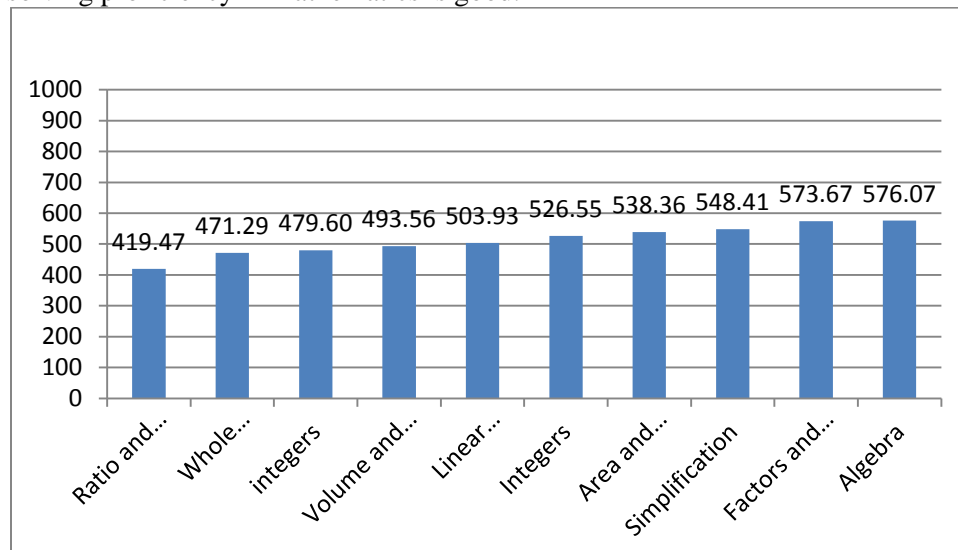


Figure 14. Content wise student's MSS

Above graph explores that ratio and proportion is the weakest while algebra is the strongest area of problem-solving.

## **Conclusions**

1. Students' mean score in Algebra is greater than any other content area while have lowest score in Ratio and Proportion.
2. Therefore students' problem-solving proficiency in the area of algebra, area and perimeter, Factors and multiples, geometry, integers, information handling, linear equations, and whole numbers is average.
3. Therefore students' problem-solving proficiency in simplification is good.
4. Therefore students' problem-solving proficiency in ratio and proportion is not good. Ratio and proportion is the weakest area of students. Three out of five questions were not solved correctly by a single student. Therefore, these three questions are above the students' ability having average difficulty.
5. Therefore students' overall problem-solving proficiency in mathematics is good.

## **Discussion**

Problem solving is a process that involves diagnosing students' cognitive processes during a problem solving activity. Testing, like in this study drives from implemented mathematics curriculum and mostly criterion referenced rather than norm referenced testing is emphasized. Usually curriculum is lengthy and bit inhuman to "cover" plenty of subject matter and therefore teachers avoid to emphasis on problem solving. In addition, due to time constraints, teachers can recognize that problem solving is advantageous for increasing understanding but may not incorporate problem solving into their tests. It is observed that Teachers often argued and gave justification for not using problem solving activities in school testing. Most of the time these arguments that it is neither measurable nor testable seems justifiable because of following reasons:

- Student strength in the classes hinders to administer the problem solving activities in the classrooms.
- Time constraint is another issue.
- Non-availability of appropriate mathematical problems.

- Shortage of problem solving material in the textbooks etc.

These assumptions are not valid. So we have to emphasize on the changes of testing practices. For this purpose we should accommodate our students by providing them invaluable mathematics experiences and to generate authentic problem solving activities. These activities can play an integral part in mathematics instruction. It must be the effort of a mathematics instructor to create that type of environment which can create conducive learning environment for problem solving.

The present study was designed by keeping in view the importance of problem solving for the cognitive development of the students. In order to meet this objective mathematics test for VI grade students was developed which only evaluate problem solving skills. This test can be used to provide feedback to all stakeholders especially the main purpose of this problem solving test is to provide feedback to teachers and students. The procedure of test development in this study involves mathematics framework, content standards and alignment of the items to these standards while using item response theory based analysis.

Mathematics disclosed unknown patterns that help us to comprehend the world around us. Mathematics as a subject believes on empiricism in which data is derived from science. Most of the mathematical reasoning is deductive which is based on inferences and proofs from natural phenomena of human behavior, and social systems.

Mathematics is not just calculation rather it is based on observations, hypothesis, testing of hypothesis and conclusions. Hence, we can say that mathematics is a science of patterns and order. Its major areas are numbers, probability, forms algorithms and change. Mostly mathematics relies on logics rather than senses. It can be concluded from above discussion that the main objective of mathematics education is to enable students to identify the association of mathematical concepts in real life. There are two basic dimensions of mathematical framework:

- Mathematical abilities
- Mathematical competencies

In the present research, in fit and outfit MS statistics was used. The power of Item difficulty and item discrimination were also used for assortment of the items for final test. Multi-dimensional analysis within items was used to validate the fit statistics. Fitness of the model was decided on the basis of the difference in deviances and estimated item parameters. The above mentioned analysis gives sound reasoning for the selection of IRT model. In National curriculum each SLO is based on standards and benchmarks .IRT provided link between item and students score. It is only possible in IRT model that students' success on an item

and the item complexity can be placed on a single scale. So the position of an item in the proficiency scale can be located. This scale gives guidance in transforming students' score into proficiencies interpretation for the skills they acquired. It means this scale enable us to provide students' ability in terms of item difficulty level. It also provides us qualitative interpretation of students' proficiencies. This proved that present problem solving test can be used for numeric and qualitative descriptions of students' proficiencies. These two descriptions can be used for different assessment purposes. These proficiencies descriptions usually depend on number of test items and their spread on the proficiency scale. In order to measure the consistency of the test it is required that the parallel test should be administered to calculate the reliability of the test. This was not the scope of the present study so the reliability dimensions of the test cannot be mathematically validated. As far as the validity is concerned external validity issues are still there in the test. In order to resolve these issues different samples of the students should be used to ensure the external validity but time constraint did not allow doing this. Different observations were taken for the validity of the test results to control internal validity issues. Relationship between item and its specific proficiency was first established by item writers and then it is empirically validated through data. Item fit statistics were used to check the validation of the extent to which this test measured the latent rate. One more step in this direction was the comparison of pilot and final test data. Same results were obtained for fit statistics.

Item-person map shows that students' ability to solve problem in mathematics is between -2 to 2. Therefore students' overall problem-solving proficiency in mathematics is good. It also shows that this test is more suitable for the students of medium ability as compared to high and low ability students.

Item format was consistent with the proficiency strands which were discussed in the literature review chapter. In order to ensure the content validity items were aligned with National Curriculum and Mathematics proficiency framework.



## References

- Alexander, P. A., White, C. S., & Daugherty, M. (1997). Analogical reasoning and early mathematics learning. In L. D. English (Ed.), *Mathematical reasoning: Analogies, metaphors, and images* (pp. 117–147). Mahwah, NJ: Erlbaum.
- Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, 23, 285–303.
- Azeem M.(2013).Development Of Math Proficiency Test Based on Item Response Theory(IRT). Unpublished Ph.D. dissertation .Division Of Education University of Education, Lahore.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press. Available: <http://books.nap.edu/catalog/6160.html>. [July 10, 2001].
- Brownell, W. A. (1935). Psychological considerations in the learning and the teaching of arithmetic. In W. D. Reeve (Ed.), *The teaching of arithmetic (Tenth Yearbook of the National Council of Teachers of Mathematics)*, pp. 1–31). New York: Columbia University, Teachers College, Bureau of Publications.
- Carpenter, T. P., Fennema, E., Fuson, K., Hiebert, J. Human, P., Murray, H., Olivier, A., & Wearne, D. (1999). Learning basic number concepts and skills as problem solving. In E. Fennema & T. A. Romberg, *Mathematics classrooms that promote understanding* (pp. 45–61). Mahwah, NJ: Erlbaum.
- Carroll, J.B.(1996).Mathematical abilities: Some results from factor analysis. In R. J. Sterburg, & T. Ben\_Zeev, (Eds.).*The nature of mathematical thinking* (pp.3-25), Mahwah, NJ: Lawrence Erlboun Associates.
- Cohen, D. K., & Ball, D. L. (2000, April). Instructional innovation: Reconsidering the story. Paper presented at the meeting of the American Educational Research Association, New Orleans.
- Davis, R. B., & Maher, C. A. (1997). How students think: The role of representations. In L. D. English (Ed.), *Mathematical reasoning: Analogies, metaphors, and images* (pp. 93– 115). Mahwah, NJ: Erlbaum.
- Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (Eds.). (1999). How people learn: Bridging research and practice. Washington, DC: National

- Academy Press. Available: <http://books.nap.edu/catalog/9457.html>. [July 10, 2001].
- Dossey, J. A., Mullis, I. V. S., Lindquist, M.M., & Chambers, D.L. (1988). The mathematics report card: Are we measuring up? Princeton NJ: Educational Testing Service.
- Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040–1048.
- English, L. D. (1997a). Analogies, metaphors, and images: Vehicles for mathematical reasoning. In L. D. English (Ed.), *Mathematical reasoning: Analogies, metaphors, and images* (pp. 3–18). Mahwah, NJ: Erlbaum.
- Enreston, S.E. (1997). Multi - component response models. In van der Linden, W.J. & Hambleton, R. R.K. (Eds.), *Handbook of modern item response theory* (pp.305-321). New York: Springer - Verlag.
- Fuson, K. C. (1992b). 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.
- Geary, D. C. (1995). Reflections of evolution and culture in children's cognition. *American Psychologist*, 50(1), 24–37.
- Gronlund, N. E. & Linn, R. L. (2005). Measurement and assessment in Teaching. New Dehli: Baba BarkhaNath Printers.
- Hambleton, R. K. Swaminathan, H. & Rogers, H. J. (1991) *Fundamentals of Item Response*
- Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*.
- Hiebert and Carpenter, 1992, Hiebert and Wearne, 1996. For work in psychology, see Baddeley, 1976; Bruner, 1960, pp. 24–25; Druckman and Bjork, 1991, pp. 30–33; Hilgard, 1957; Katona, 1940; Mayer, 1990; Wertheimer, 1959. Theory. London. Sage Publications.
- Hiebert, J. (Ed.). 1986). Conceptual and procedural knowledge: The case of mathematics. Hillsdale, NJ: Erlbaum
- Hiebert, J. & Wearne, D. (1996). Instruction, understanding, and skill in multi digital addition and subtraction. *Cognition and Instruction*, 14, 251–283.

- Kilpatrick, J. (1985). Doing mathematics without understanding it: A commentary on Higbee and Kunihiro. *Educational Psychologist*, 20(2), 65–68.
- Lai, C.P. (1998). Statistical modeling of student's performance in multi-component tasks. Unpublished Ph.D. Thesis, Department of Learning and Educational Development, University of Melbourne
- Li min, WU,M.(2003)The Application of Item Response Theory to Measure Problem-solving Proficiencies. Unpublished Ph.D. dissertation .Department of Learning and Educational Development The University of Melbourne.
- Maher, C. A., & Martino, A. M. (1996). The development of the idea of mathematical proof: A 5-year case study. *Journal for Research in Mathematics Education*, 27, 194–214.
- Mayer, R.E. (1992).*Thinking, problem-solving, cognitive.*(2nd Ed.) New York. Freeman and Company.
- Mayer, R.E., & Hegarty, M.(1996).The process of understanding mathematical problems. In R.J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp.29- 53).Mahwah, NJ: Lawrence Erlbaum Associates. National Council of Teachers of Mathematics (1989).Curriculum and evaluation standards for school mathematics. Reston, VA: NCTM.
- Pesek, D. D., & Kirshner, D. (2000). Interference of instrumental instruction in subsequent relational learning . *Journal for Research in Mathematics Education*, 31, 524–540.
- Pólya, G. (1945). How to solve it: A new aspect of mathematical method. Princeton, NJ: Princeton University Press
- Reckase, M.D.(1997).A linear logistic multi-dimensional model for dichotomous item response data. In W. J. Van der Linden & R. K. Hambleton (Eds.) *Handbook of modern item response theory* (pp.271-286).New York:
- Resnick, L. B. (1987). Education and learning to think. Washington, DC: National Academy Press.
- Suen, H. K. (1990) Principles of Test Theories . Pennsylvania State University. Hove and London. Lawrence Erlbaum Associates, Publishers.
- Schoenfeld, A. H. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education*, 20, 33

- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57, 1–22.
- Webb, N. L. (1997b, January). Determining alignment of expectations and assessment in mathematics and science education.
- Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transaction*, 8:3 p.370 Retrieved from <http://www.rasch.org/rmt/rmt83b.htm> 24.08.2013.
- Yaffee, L. (1999). Highlights of related research. In D. Schifter, V. Bastable, & S. J. Russell with L. Yaffee, J. B. Lester, & S. Cohen, *Number and operations: Making meaning for operations. Casebook* (pp. 127–149). Parsippany, NJ:

---

*Received on: 02.07.2015*

*Revised on: 09.06.2016*