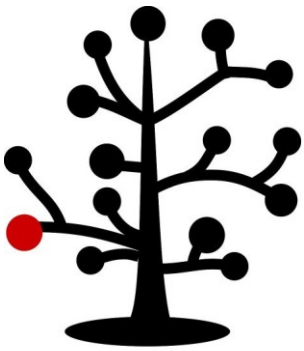


N.i.D.S.

NATIONAL INCOME DYNAMICS STUDY

Numeric Literacy in South Africa: Report on the NIDS Wave 1 Numeracy Test

Technical Paper no. 5

Patrick Griffin
Assessment Research Centre
p.griffin@unimelb.edu.au

Murray Leibbrandt,
Southern African Labour & Development Research Unit
Murray.Leibbrandt@uct.ac.za

Masa Pavlovic
Assessment Research Centre
masap@unimelb.edu.au

Tia Linda Zuze
SALDRU Research Affiliate
tlzuze@gmail.com

November 2010

Contents

1. Introduction

The National Income Dynamics Survey is the first study of its kind to provide accurate information about the level of basic mathematics literacy among South Africans aged between 12 and 72. This report provides a first look at the content and design of the mathematics test and presents a general overview of the findings on mathematics literacy. There were several reasons why the literacy skills of the population were particularly relevant to the NIDS framework. Aside from the personal barriers faced by individuals with limited numeracy ability, there are national costs related to the country's economic growth and development that go beyond the individual.

South Africa's complex educational history has left many individuals without the necessary quantitative skills to function in society. Adult literacy for South Africans is estimated to be 88% for the period 2000 to 2006 (UNESCO 2008). And yet study after study reveals that South Africans of various ages lag desolately behind their regional and international peers in terms of mathematical skills (Department of Education 2005; Moloi and Strauss 2005; Mullins et al. 2004). Even when assessment has been based on local benchmarks, the numeric ability levels remain below prescribed standards (Department of Education 2005). Previous assessments have concentrated on a specific age group or educational level. The numeracy module is unique in determining national numeric ability across a wide age range and among respondents that have been exposed to various phases of South African education. Formal schooling is by no means the only route to acquiring knowledge. An added contribution of this module is to determine where mathematical competency exceeds expectations. The report is structured as follows. In Section 2, we discuss the steps followed in constructing the numeracy assessment. This section includes detailed information about the scope of the test, the piloting exercise and the item analysis of the final results. Sections 3 and 4 deal with the important issues of the scale reliability and its validity. A detailed description of the skills audit that was undertaken to derive competency levels from the numeracy assessment is also provided. The competency levels are useful because they link a respondent's score to the ability to carry out specific numerical tasks.

2. Constructing the Numeracy Assessment

The ten-minute test was administered to eligible household members as part of the scheduled household data collection exercise. The test design covered five key mathematical domains: numeration, measurement, algebra, space and data. To ensure consistency across tests, field workers were instructed to provide as limited verbal prompts as possible during the assessment. The diversity of educational experiences among respondents presented a unique challenge for the test construction. To overcome this problem, five different testing instruments were made available, corresponding to different age or education levels. Common items were included across tests to ensure that responses could later be combined into a single scale of mathematical literacy.

Test Content

Five tests representing five different age or education levels were constructed. Each test consisted of 15 items and covered five different domains with three items per domain. There were five overlapping items between levels of each test. Each of the five overlapping items belonged to a different domain.

The dimensions were defined as:

- ***Numeration****:* Operations with numbers, rounding, place values, fractions, decimals, percentages, and ratios.
- ***Algebra****:* Number sequences, geometric patterns, algebraic expressions, function and function values.
- ***Measurement****:* Conversion between appropriate time, money, length, and volume units, relationship between surface area and volume of an object.
- ***Space****:* Geometric shapes, basic geometry.
- ***Data:*** Table and graphical representation of data, data manipulation, probability of an independent and compound event.

The first step in test construction involved analysing the most recent version of the mathematics curriculum to identify the most basic elements to be assessed by each of the five tests. Therefore, each level covered fundamental mathematics skills considered to be taught at the following educational levels.

1. No Education to Grade 3;
2. Grade 3 to Grade 6 ;
3. Grade 7 to Grade 9;
4. Grade 10 to Grade 12;
5. Higher Education.

Items were designed to be as straightforward as possible and were written to assess mathematical literacy on five levels of ability and five mathematical dimensions. The tests were not intended to demonstrate ability levels in individual dimensions. Such an assessment would have required a testing instrument that would have been too long and complex to administer in the limited time available to field workers.

It is worth noting here that although the initial test design contained five unique testing instruments, the fifth option (higher education) proved to be too difficult for respondents and was subsequently dropped from the design. Therefore the results reported here exclude descriptions of the higher education test. In addition, missing items were treated as incorrect. Because of the length of the test, the assumption was that the reason for not providing an answer was because respondents did not know the correct answer.

Table 1 provides details of each dimension for all ability levels (including level 5). For example, requirements for the numeration dimension ranged from simple arithmetic manipulation at the lower levels to working with fractions, ratios and percentages at higher levels. Similarly, the range for the geometry dimension stretched from identifying two dimensional shapes at the lowest level to calculating volumes and proportions at higher levels.

Table 1: Test blueprint and proposed levels

Level	Aimed at years of education	Numeration	Measurement	Geometry	Data	Algebra
Level 1	No education to Year 3	Solves verbally or graphically stated additions	Simple money operations .Compare prices(1)	Recognises and names two dimensional shapes.	Sort physical object according to one property i.e. colour shape.	Completes number sequence.
		Division with single digit numbers	Conversion between standard length units(1)	Recognises and names three dimensional shapes.	Sort/organise data according to numerical attribute.	Completes number sequence. Solve simple algebraic expressions.
		Basic arithmetic manipulation with 2 digit numbers namely addition subtraction multiplication and division.	Conversion between standard time, length, and volume units.	Recognises two dimensional shapes on the faces of tree dimensional object	Able to read and interpret data presented in simple table or list	Completes number sequence. Solve simple algebraic expression Recognises numeric and geometric patterns.
Level 2	Year 3 to Year 6	Simple operations with integers. BODMAS. Recognises place value of digits in whole 3 digit numbers	Estimate weight of real life object.	Finds relationship between perimeter and area.	Organise and examines numerical data	Completes more complex number sequences
		Calculations with integers, common fractions and decimals, recognise place values in 3 digit numbers.	Calculate time. Conversion between appropriate time units	Relationship between perimeter area and volume.	Organise and examines numerical data.	Solve simple algebraic expressions
Level 3	Year 7 to Year 9	Multiple operations with whole numbers with or without brackets. Knowledge of prime numbers	Solve problems involving calculations and conversion between appropriate time units S.I. units. Calculates, perimeter, area and volume of an object.	Uses geometry of straight lines and triangles.	Organizes and describes ungrouped numerical data by determining mean. Read data from bar and double bar graphs	Finds the function from the table values
		Multiple operations with whole numbers with or without brackets. Knowledge of fractions, decimals, and percentages. Solve problems	Solve problems involving calculations and conversion between appropriate time units. Knowledge of relationship	Calculates distance between two points in the co-ordinate system .Familiarity with rotations, reflections,	Understands probability of an event .Reads data presented in a variety of ways. Identifies the largest	Uses formulas and equations to solve a problem. Simplify algebraic expression

Level	Aimed at years of education	Numeration	Measurement	Geometry	Data	Algebra
Level 4	Year 9 to Year 12	that involve comparing two different quantities (ratio).	between surface area and perimeter of rectangles and squares.	translations and symmetry properties of the object.	and smallest scores in the data set.	
		Solves problems that involve ratio, rate, and proportion.	Knowledge of relationship between surface area and perimeter of rectangles and squares.	Recognises axes of symmetry and calculates volumes of 3D objects	Determine probability of an event	Solves problems that involve ratios. Solves linear equations. Uses factorisation to simplify algebraic expressions
		Uses a range of techniques and tools to perform calculations efficiently. Absolute value of decimal, rational and irrational numbers	Converts between mm2 cm2 m2 and km2 Use of Pythagoras theorem to solve problems involving missing length in known geometric figures and solids.	Volumes and proportions.	Displays and reads data from bar graphs and double bar graphs, histogram and pie charts.	Uses difference of squares and cubes to simplify algebraic expression.
Level 5	Higher education	Knowledge of exponents and logarithms.	Understands the effect of multiplying one dimension by constant factor in 3D object on volume	Use the formula for surface area and volume of pyramids, spheres ,cones and cylinders	Displays and reads data from bar graphs and double bar graphs, histogram or pie charts. Determine the probability for the compound events.	Uses the laws of exponents and factorisation to simplify and solve equations. Determines the equation or formula from given graphs.
		Simplify expressions using the law of exponents.	Knowledge of trigonometry, sin cos …	Use the formula for surface area and volume of pyramids, spheres ,cones and cylinders	Determines the probability of compound events and recognises dependent and independent events.	Simplify expression involving logarithms

Draft items were reviewed by a panel at the University of Melbourne's Assessment Research Centre (ARC). In reviewing the items, panel members considered whether the knowledge required to answer an item correctly corresponded to the strand and education level targeted.

The Pilot Test

A pilot exercise was undertaken to identify poorly performing test items. Questions could be dropped if they were unsuitable for a given ability level. In some instances, the questions were unclear and were deemed inappropriate. The pilot was also used to determine which version of the test was most suitable to administer to a particular household member. As a result of the trials, several items were adjusted and some were omitted from the test forms entirely.

Test Calibration

Item response modelling was used to develop a single scale of mathematical literacy. The underlying assumption behind this model is that the probability of responding correctly to a question is simultaneously dependent on the difficulty of the question and the inherent ability of the individual taking the test. The ability and difficulty parameters are mapped onto a single interval scale in a process called calibration. Both parameters are measured in the same units (often referred to as logits). The existence of a single scale enables both persons and items to be placed on the same continuum. The estimates are then plotted onto a chart called a variable map which illustrates the relative position of respondents against the difficulty levels assigned to each of the test items. Concurrent estimation procedures were used to equate the four tests in question. All 46 test items were simultaneously calibrated, irrespective of the test level or the population age group assessed.

Item Analysis

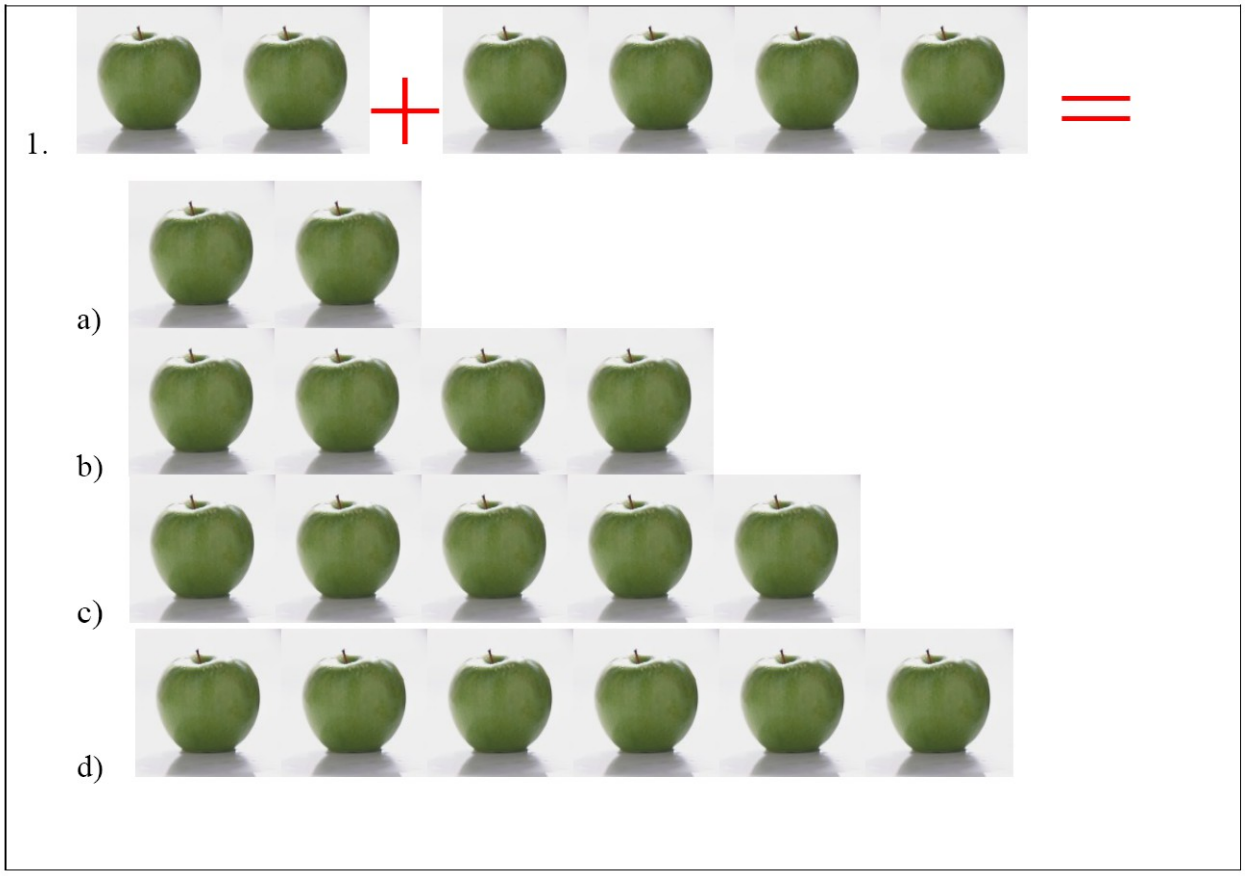
One of the goals of item response modelling is to develop a test that is both accurate and valid. By this we mean that the results are reliable and would remain consistent should similar tests be administered to the same group of people. It also implies that the performance of respondents relative to each other would remain constant. In the past, raw scores have been used to evaluate ability.. Tests are administered and the results are simply summed for each individual. The drawback of using raw scores is that although they provide an accurate ordering of results (a candidate who scores 10 has outperformed a candidate who has scored 8), the distance between scores is fairly arbitrary. For example if a test is very easy and respondents of varying ability perform well, then the raw scale will fail to capture the magnitude of their ability differences. An additional advantage of an item scale over a raw score is that an item scale score

can be linked to proficiency in specific tasks. Thus, if person 'X' scores 15, it is possible to relate this score to a level of numeric competency.

A classical item analysis was performed on each of the questions to verify the item characteristics and to test whether the distractor items were functioning properly. The analysis for item 1 is shown in Figure 1 below. The analysis includes several useful pieces of information. The first is the number of people who responded to the item - in this case 473. The discrimination index measures how well an item separates respondents based on their ability relative to other items of similar difficulty. The value of 0.25 is relatively low, but this is possibly due to the fact that 87.53 percent of the persons attempting the item selected the correct answer. The item difficulty was -2.70 which represent a value of 2.7 logits below the mean item difficulty, implying that this was a relatively very easy item.

Two of the alternatives for this item did not work well. For example alternatives 'a' and 'c' attracted less than 1 percent of the responses. Nevertheless an entry level item should always be used in the introductory section of a test.

Figure 1: Example of Item Analysis



Item 1

Cases for this item 473 Discrimination 0.25

Item Threshold(s): -2.70 Weighted MNSQ 1.03

Item Delta(s): -2.70

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
a	0.00	2	0.42	-0.11	-2.44(.015)	-2.36	0.17
b	0.00	10	2.11	-0.14	-3.02(.003)	-1.40	0.89
c	0.00	4	0.85	-0.02	-0.48(.628)	-1.41	1.39
d	1.00	414	87.53	0.25	5.49(.000)	-0.38	0.84
z	0.00	43	9.09	-0.18	-3.99(.000)	-1.19	1.17

The weighted mean square residual (INFIT) estimates the extent to which the data for this item are consistent with the expected performance under the Rasch model (Wright and Masters 1982). These estimates are calculated by weighting the variance of the item response with the squared value of the standardised residual. The output in Figure 1 above shows a value of 1.03 against an expected value of 1.00 and an expected range of +/- 0.3. Hence the item qualifies for retention in the scale. The plausible value average (PV1Avg) for the correct answer is higher than for each of the incorrect alternatives, which supports the assertion that the more capable respondents chose the correct alternative and the less able respondents chose the incorrect alternatives. Repeating this analysis for all of the items highlighted that there were some problematic items which needed attention.

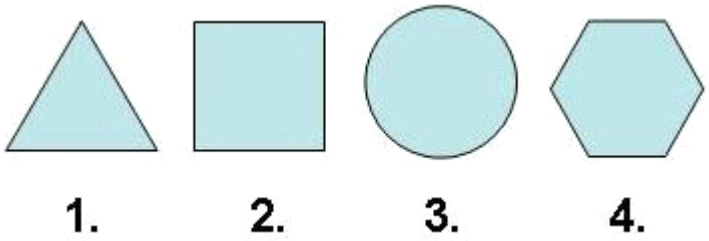
The measurement errors were relatively small, but this is expected given the large sample size. All but one of the INFIT values were within the range of 0.9 to 1.1, and hence there was sound evidence of a dominant underlying dimension in the variable being measured. The mean item difficulty was arbitrarily set to be zero. The variance of item difficulty values was 1.6 with a reliability of item separation of 0.97. The mean item INFIT was 0.999 with a variance of 0.003. There were no items with zero scores and no items with perfect scores. The mean participant ability estimate was -0.5176, indicating that the respondents' ability level was slightly lower than the difficulty of the overall test. The variance of the pupil ability estimates was 1.19 which was lower than the variance of item difficulties. This suggested that item difficulties were more spread out than the ability range of the tested population.

Several items had to be removed from the final analyses of the data.

Item 'w1_numq6_w2numq6' was altered between the trial and final assessment and presented two possible correct answers after the circle was altered.

Figure 2: First Example of a Corrected Item

1. Which shape is the odd one out?



a) 1.
b) 2.
c) 3
d) 4.

Item 6

---

item:6 ('w1_numq6_w2numq6')

Cases for this item 1601 Discrimination 0.18

Item Threshold(s): 0.41 Weighted MNSQ 1.04

Item Delta(s): 0.41

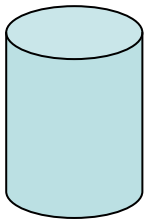
Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
a	0.00	172	10.74	0.05	1.86(.064)	-0.56	0.69
b	0.00	95	5.93	-0.05	-1.91(.056)	-0.82	0.64
c	1.00	475	29.67	0.18	7.23(.000)	-0.26	0.76
d	0.00	632	39.48	0.13	5.30(.000)	-0.55	0.77
z	0.00	227	14.18	-0.43	-18.82(.000)	-1.37	0.86

Alternatives ‘a’, ‘c’ and ‘d’ were all popular and 14% of respondents did not select any alternative. The change was to remove the ‘smiley face’ from the circle, leaving item with multiple interpretations.

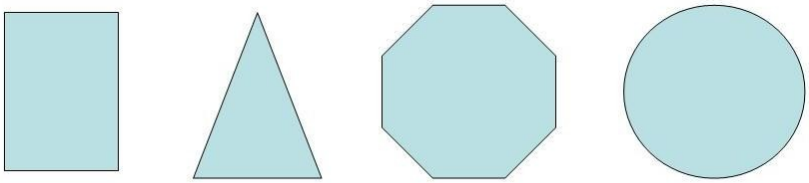
On test level 2 the following item was omitted:

Figure 3: Second Example of a Corrected Item

2. If this cylinder was cut using a vertical plane



What would be the shape of the cut surface?



a) b) c) d)

item:23 ('w2_numq12')

Cases for this item 1128 Discrimination 0.13

Item Threshold(s): 0.78 Weighted MNSQ 1.06

Item Delta(s): 0.78

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
a	1.00	252	22.34	0.13	4.38(.000)	-0.34	0.78
b	0.00	109	9.66	-0.06	-1.90(.058)	-0.78	0.64
c	0.00	140	12.41	-0.06	-2.17(.031)	-0.76	0.66
d	0.00	469	41.58	0.27	9.47(.000)	-0.49	0.71
z	0.00	158	14.01	-0.43	-16.05(.000)	-1.38	0.78

It can be seen that there is a problem in identifying the vertical and horizontal planes.

Figure 4 provides an example of a question that had to be omitted because of a typographical error in the final version of the printed test.

Figure 4: An Example of an Omitted Item

3. $\dfrac{1+\frac{1}{2}}{1-\frac{1}{2}} - \dfrac{1-\frac{1}{2}}{1+\frac{1}{2}} = ?$

a) -1
b) 0
c) 1
d) 8/3

item:37 ('w4_numq2')
Cases for this item 1220 Discrimination -0.07
Item Threshold(s): 1.71 Weighted MNSQ 1.04
Item Delta(s): 1.71

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
a	0.00	255	20.90	-0.07	-2.35(.019)	-0.68	0.64
b	0.00	304	24.92	0.07	2.32(.021)	-0.50	0.65
c	0.00	440	36.07	0.12	4.22(.000)	-0.46	0.63
d	1.00	137	11.23	-0.07	-2.44(.015)	-0.37	0.61
z	0.00	84	6.89	-0.15	-5.15(.000)	-0.84	0.83

In this item the minus sign was missing, thus rendering the item meaningless. The data corresponding to these test items were removed from the analysis. The negative discrimination and the roughly equal ability of each respondent group indicate guessing. Without the minus sign the item could not be interpreted.

The remaining 44 items were used in a concurrent equating exercise to estimate the ability parameters of the respondents to the numeracy survey. The calibration of the test was therefore conducted using only the 44 retained items.

Details of these analyses are provided in Table 2. The following descriptive statistics are included in the table:

1. The item difficulty logit;
2. The measurement error (SEM);
3. The outfit estimate (unweighted Mean Square residual);
4. The infit estimate (Weighted Mean Square residual);
5. The proportions of respondents selecting each alternative distracter and the proportion of participants omitting the item;
6. The number of respondents who attempted the item;
7. The proportion of the population that obtained the correct response;
8. The discrimination index of each item.

Blank data rows indicate omitted items.

The first column in Table 2 lists the item's position in the data file. The second column identifies the item label and the third column lists the item variable name. The coding of the variable name reveals the level of the test and the number within each level. For example, item variable name w1_numq2 indicates that the item belongs to level 1 and is test item 2. In the case of common items such as w1_numq14_w2numq13, the variable name indicates that the item can be found in level one item 14 and level two test item 13.

The summary statistics for each item were as follows:

Table 2: Item Analysis Summary

----	Item Label	Variable Name	Logit	SD	Outfit	Infit	A %	B %	C %	D %	Miss%	N	p	Discrimination
1	Addition with single digit numbers	w1_numq1	-2.678	0.092	1.16	1.03	0.42	2.11	0.85	87.53	9.09	473	0.88	0.24
2	Division with single digit number	w1_numq2	-1.962	0.085	0.9	0.91	3.38	4.02	79.07	5.29	8.25	473	0.79	0.44
3	Multiplication of one digit numbers	w1_numq3_w2numq3	-2.52	0.062	0.88	0.93	0.94	2.44	5.25	84.76	6.62	473	0.85	0.31
4	Extends simple patterns involving numbers	w1_numq4	-1.295	0.08	0.91	0.93	10.78	5.5	68.29	4.02	11.42	473	0.68	0.43
5	Recognise relationship between the numbers. Reciprocal relationship between division and multiplication.	w1_numq5	-0.392	0.076	0.88	0.91	6.55	12.9	12.26	50.32	17.97	473	0.5	0.45
6	Identifies geometric patterns	w1_numq6_w2numq6												
7	Simple money operations. Compare prices.	w1_numq7	-1.903	0.084	0.71	0.82	2.96	3.38	78.22	3.38	12.05	473	0.78	0.59
8	Reads time in terms of hours and minutes.	w1_numq8	-1.488	0.081	0.97	0.95	1.9	71.67	2.96	13.74	9.73	473	0.72	0.38
9	Simple length conversion (cm to mm)	w1_numq9_w2numq9	0.409	0.052	0.94	0.96	30.73	29.92	11.06	12.62	15.68	1601	0.31	0.33
10	Surface area estimation	w1_numq10	-1.845	0.084	0.79	0.86	6.34	2.33	3.17	77.38	10.78	473	0.77	0.52
11	**Counting sides in 3D objects**	**w1_numq11_w2numq10**	**-0.123**	**0.05**	**0.92**	**0.95**	**36.04**	**8.81**	**41.22**	**4.56**	**9.37**	**1601**	**0.41**	**0.42**
12	Recognises two dimensional shapes on the faces of three dimensional objects	w1_numq12	1.129	0.084	1.15	1.04	57.51	5.29	21.56	5.29	10.36	4.73	0.22	0.17
13	Groups data according to geometrical attribute	w1_numq13	0.461	0.079	0.93	0.96	2.33	44.19	5.29	32.98	15.22	473	0.33	0.33
14	Reads and interprets data presented in table	w1_numq14_w2numq13	-1.466	0.052	0.9	0.92	68.83	7.81	4.93	5.31	13.12	1601	0.69	0.37
15	Groups data according to numerical attribute	w1_numq15	0.139	0.077	0.92	0.95	39.32	6.34	8.67	29.81	15.86	473	0.39	0.37
16	BODMAS	w2_numq1	0.276	0.059	1.1	1.06	11.88	11.35	31.83	34.66	10.28	1128	0.32	0.19
17	Place value in 4 digit numbers	w2_numq2_w3numq2	-0.646	0.039	1.01	1.01	10.83	15.41	51.99	12.67	9.1	2660	0.52	0.26
18	Solve simple equations	w2_numq4_w3numq4	-2.048	0.046	0.93	0.95	5.68	3.61	78.87	6.24	5.6	2660	0.79	0.35

----	Item Label	Variable Name	Logit	SD	Outfit	Infit	A %	B %	C %	D %	Miss%	N	p	Discrimination
19	Recognise and extends number patterns.	w2_numq5	-0.621	0.056	0.95	0.96	32.45	50.35	4.88	6.47	5.85	1128	0.5	0.37
20	Reads digital clock time in terms of hours an minutes.Calculates the time difference	w2_numq7_w3numq7	-0.62	0.039	0.94	0.95	16.99	11.2	51.43	10.23	10.15	2660	0.51	0.31
21	Correct estimation of weight	w2_numq8	-0.861	0.057	0.9	0.92	0.98	55.5	1.86	1.15	40.51	1128	0.56	0.42
22	Relationship between sides and perimetr of a 2D object.	w2_numq11_w3numq10	0.157	0.041	0.93	0.95	36.35	13.57	5.6	34.85	9.62	2660	0.35	0.29
23	Recognises the cut surface of the 3D objects	w2_numq12												
24	Calculates the mean	w2_numq14_w3numq14	1.289	0.05	1.1	1.04	27.41	16.17	21.05	21.62	13.76	2660	0.16	0.11
25	Reads and interprets the information presented in a bar chart	w2_numq15	-0.318	0.057	0.92	0.94	17.82	43.88	4.61	20.92	12.77	1128	0.44	0.39
26	Fractions and order of operations(BODMAS)	w3_numq1	1.403	0.063	1.16	1.05	40.27	21.87	14.95	11.95	10.97	1532	0.15	0.05
27	Calculating percentages	w3_numq3_w4numq3	0.47	0.041	1.03	1.02	21.58	19.3	19.8	29.47	9.85	2752	0.3	0.19
28	Recognise and extends number patterns.	w3_numq5	0.556	0.053	0.97	0.99	26.63	26.83	9.66	27.68	9.2	1532	0.28	0.23
29	Recognises mathematical relationship between values	w3_numq6	0.334	0.052	1	1	26.65	19.32	31.85	11.68	11.49	1532	0.32	0.21
30	Converts m2 into cm2	w3_numq8	2.554	0.082	1.19	1.01	38.45	27.61	17.3	5.55	11.1	1532	0.06	0.05
31	Convert days into minutes	w3_numq9_w4numq8	1.453	0.063	0.9	0.95	35.7	24.02	12.92	14.36	12.99	1532	0.14	0.27
32	Relationship between volume and surface in 3D objects	w3_numq11_w4numq10	0.405	0.041	1.04	1.04	32.81	9.88	17.22	30.7	9.38	2752	0.31	0.13
33	Calculates distance between two points in the Cartesian co-ordinate system	w3_numq12	1.517	0.064	1.17	1.05	33.09	30.09	13.64	12.6	10.57	1532	0.14	0.02
34	Understands probability of an independent event.	w3_numq13_w4numq14	0.994	0.045	1.04	1.03	12.43	37.65	16.93	20.57	12.43	2752	0.21	0.12
35	Reads and interprets data from double bar graphs	w3_numq15	0.887	0.056	1.2	1.07	22.06	3.79	9.46	56.14	8.55	1532	0.22	0.04
36	Calculations involving exponents of 2	w4_numq1	-0.47	0.054	1.03	1.03	25.08	4.75	49.67	16.39	4.1	1220	0.5	0.17

----	Item Label	Variable Name	Logit	SD	Outfit	Infit	A %	B %	C %	D %	Miss%	N	p	Discrimination
37	Multiple calculations with fractions	w4_numq2												
38	Solving mathematical equations	w4_numq4	0.002	0.055	1.04	1.03	18.85	34.75	39.26	3.03	4.1	1220	0.16	0.16
39	Simplify algebraic expression using the expression for difference of two squares	w4_numq5	0.936	0.062	1.11	1.04	17.7	28.36	21.64	23.61	8.69	1220	0.22	0.09
40	Recognise and extends number patterns.	w4_numq6	0.3511	0.057	1.05	1.04	21.07	32.7	32.05	9.34	4.84	1220	0.32	0.13
41	Triangle.Sum of angles. Ratios	w4_numq7	0.38	0.057	1.05	1.03	21.89	31.48	22.38	14.51	9.75	1220	0.31	0.13
42	Convert days into hours	w4_numq8	0.615	0.059	1	0.99	31.72	23.77	27.05	12.38	5.08	1220	0.27	0.22
43	Relationship between volume and capacity	w4_numq9	1.577	0.07	1.01	1	29.51	27.62	21.39	13.2	8.28	1220	0.13	0.15
44	Calculates volume of a cone	w4_numq11	1.517	0.069	1.12	1.04	13.85	31.23	22.79	22.79	9.34	1220	0.14	0.04
45	Axes of symmetry	w4_numq12	-0.576	0.054	1.12	1.09	10.33	28.93	52.05	3.93	4.75	1220	0.52	0.05
46	Knows the definition of median	w4_numq13	-0.364	0.054	1.1	1.08	47.3	20.08	15.49	9.75	7.38	1220	0.47	0.07
47	**Reads informations from stacked bar graphs.**	**w4_numq15**	**2.387**	**0.412**	1	0.98	73.93	8.77	3.03	6.56	7.7	1220	0.16	0.16

3. The Variable Map

Concurrent equating enables all four tests used in the study to be linked onto the same underlying scale. After the problematic items were omitted, the remaining 44 items were analysed and merged into a single data file. The calibration enables all respondents to be mapped onto the same continuum of 44 items as if each person responded to all items. The variable map is presented in Figure 5 below and calibration statistics are presented in Table 3.

There are several characteristics of the variable map. The 'x' characters on the left of the figure represent respondents. Each 'x' represents approximately 28 persons. A total of 4,353 respondents answered the four tests. On the right hand side is a series of numbers representing the item order on the data file and the codes used in Table 3.

Figure 5: The Variable Map

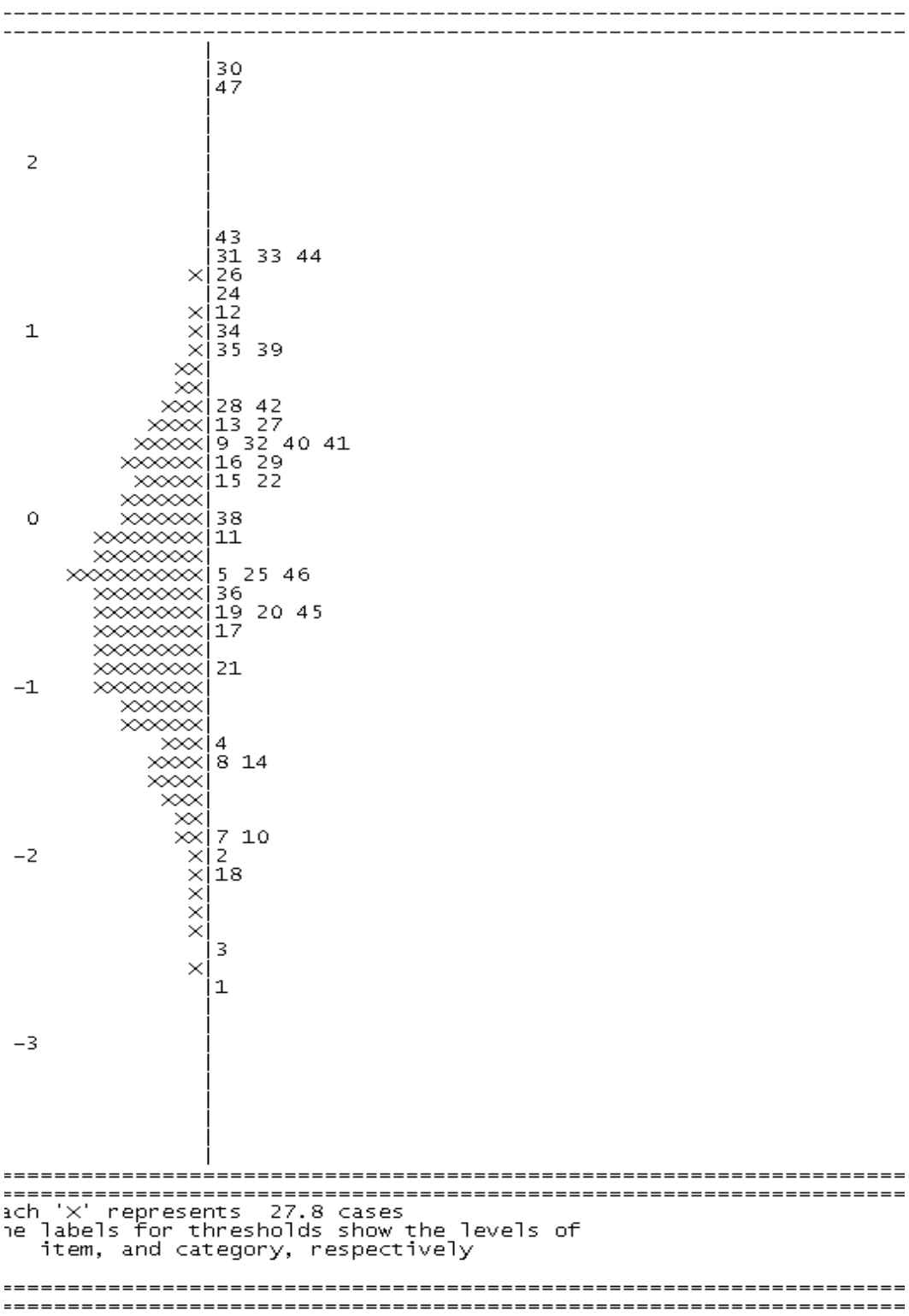
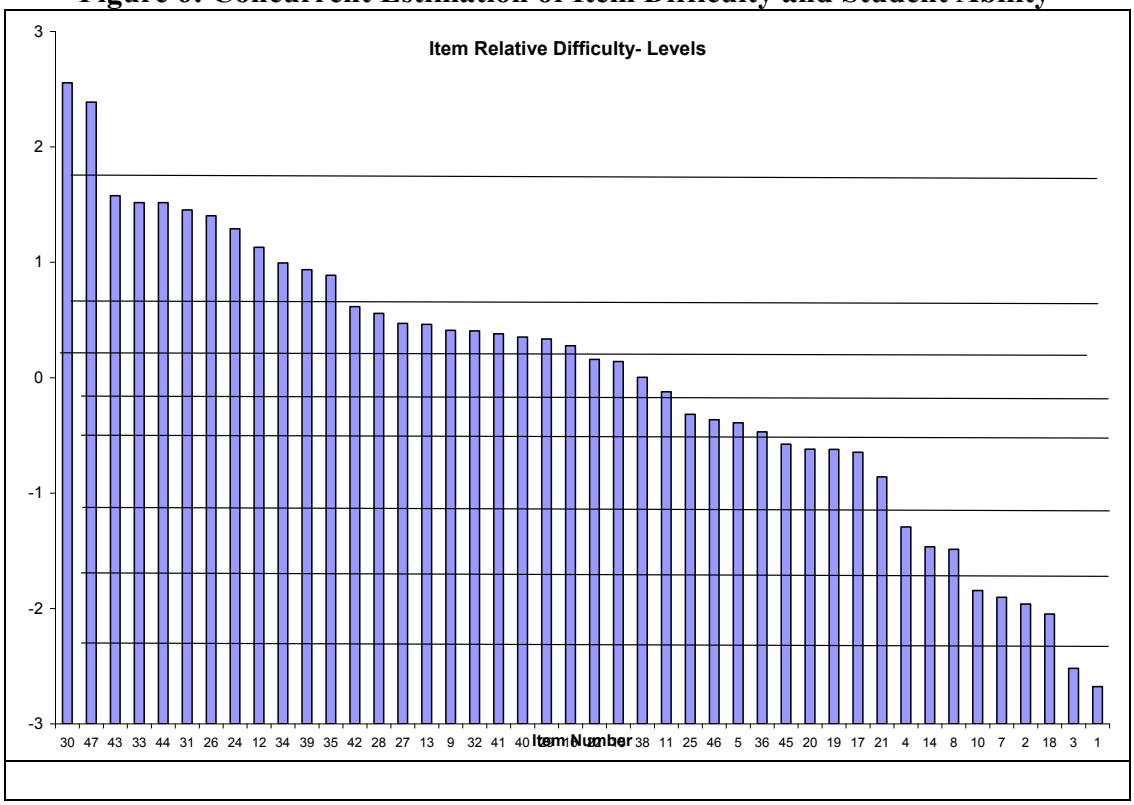


Figure 6: Concurrent Estimation of Item Difficulty and Student Ability

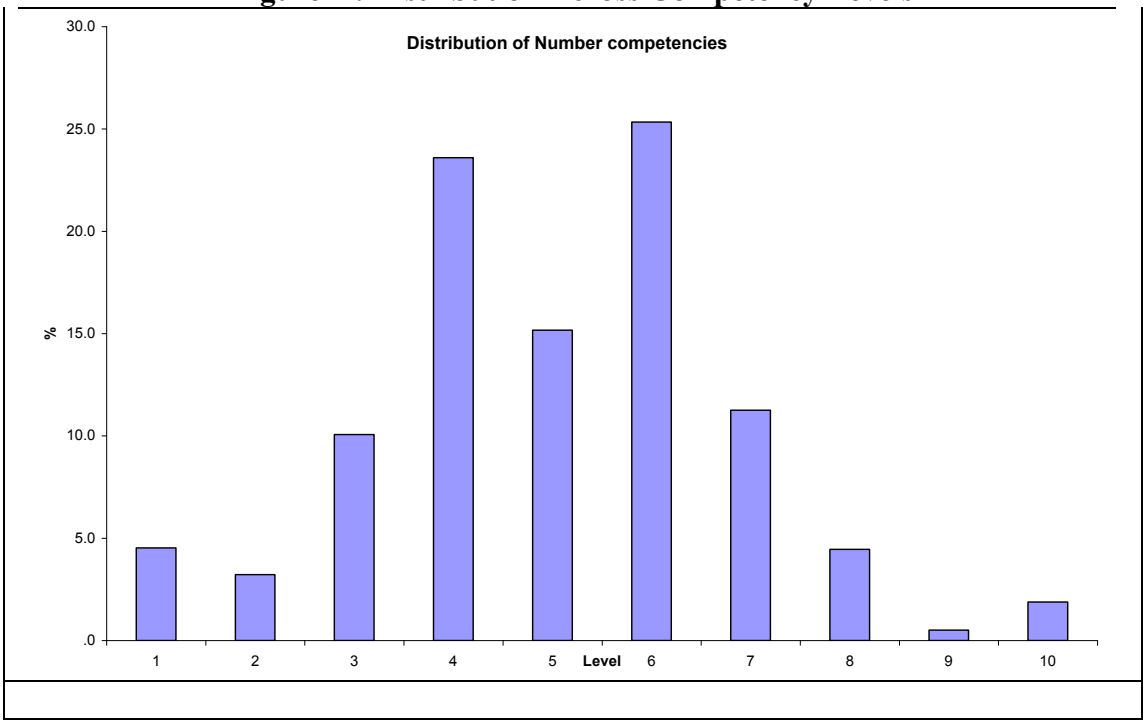


Sorting the 44 items by their relative difficulty (Figure 6 above) enables the identification of 10 levels of competency (shown in Table 3 and Figure 7 below). The competency levels are defined by the skills audit of the items within each level. Thresholds between levels are identified from the plot in Figure 6 where there is a discernable rise of difficulty separating the item set. These align roughly with the item clusters in the variable map in Figure 5.

Table 3 Possible Numeracy Competency Levels

Level	Item	Item Label	Logit
10	30	Converts m2 into cm2	2.554
10	**47**	**Reads information from stacked bar graphs.**	**2.387**
9	43	Relationship between volume and capacity	1.577
9	33	Calculates distance between two points in the Cartesian co-ordinate system	1.517
9	44	Calculates volume of a cone	1.517
9	31	Convert days into minutes	1.453
8	26	Fractions and order of operations(BODMAS)	1.403
8	24	Calculates the mean	1.289
8	12	Recognises two dimensional shapes on the faces of three dimensional objects	1.129
8	34	Understands probability of an independent event.	0.994
8	39	Simplify algebraic expression using the expression for difference of two squares	0.936
7	35	Reads and interprets data from double bar graphs	0.887
7	42	Convert days into hours	0.615
7	28	Recognise and extends number patterns.	0.556
7	27	Calculating percentages	0.47
7	13	Groups data according to geometrical attribute	0.461
7	9	Simple length conversion (cm to mm)	0.409
7	32	Relationship between volume and surface in 3D objects	0.405
7	41	Triangle.Sum of angles. Ratios	0.38
7	40	Recognise and extends number patterns.	0.3511
7	29	Recognises mathematical relationship between values	0.334
6	16	BODMAS	0.276
6	22	Relationship between sides and perimeter of a 2D object.	0.157
6	15	Groups data according to numerical attribute	0.139
6	38	Solving mathematical equations	0.002
6	**11**	**Counting sides in 3D objects**	**-0.123**
5	25	Reads and interprets the information presented in a bar chart	-0.318
5	46	Knows the definition of median	-0.364
5	5	Recognise relationship between the numbers. Reciprocal relationship between division and multiplication.	-0.392
5	36	Calculations involving exponents of 2	-0.47
4	45	Axes of symmetry	-0.576
4	20	Reads digital clock time in terms of hours an minutes. Calculates the time difference	-0.62
4	19	Recognise and extends number patterns.	-0.621
4	17	Place value in 4 digit numbers	-0.646
4	21	Correct estimation of weight	-0.861
3	4	Extends simple patterns involving numbers	-1.295
3	14	Reads and interprets data presented in table	-1.466
3	8	Reads time in terms of hours and minutes.	-1.488
3	10	Surface area estimation	-1.845
2	7	Simple money operations. Compare prices.	-1.903
2	2	Division with single digit number	-1.962
2	18	Solve simple equations	-2.048
1	3	Multiplication of one digit numbers	-2.52
1	1	Addition with single digit numbers depicted as objects	-2.678

Figure 7 : Distribution Across Competency Levels



4. Validity

The consistency between what the numeracy scale is intended to measure (as set out in the specifications) and the evidence that can be collected to demonstrate the existence of the construct, is called the nomological validity[1]. In this study it was hypothesised that a level of education would correctly guide field workers as to the test level most appropriate to the respondent Differences were expected across the four groups of respondents that answered the four different tests. Similarly, one would not expect large differences between respondents who had attended similar levels of education, who differed in gender, or had a different home location within the same test.

These relationships to external characteristics of the respondents by the construct give us further evidence of its validity. By and large these assumptions have been supported by the data. Wright and Masters (1982) have also argued that the construct is defined by the skills needed to select the correct response and the manner in which these skills are distributed across the scale. If the items and their attendant skills are sufficiently separated along the scale, then the interpretation of the construct is enhanced. This is taken as evidence of construct validity. In a similar way, the extent to which the respondents can be separated along the variable scale based on theoretical ability estimates is taken to be evidence of criterion validity. The item separation index was 0.99 and the reliability of the person separation index was 0.613. The mean squared INFIT index was 0.99 with a variance of 0.0032. Put together, the evidence indicates that the test was measuring a single dominant variable and that a single dominant latent variable underpinned the set of items. Therefore there is sufficient support of validity across the testing instruments.

[1] A form of construct validity, it is the degree to which a construct behaves as it should within a system of related constructs called a nomological set.

5. Conclusion

The NIDS numeracy score is based on a ten minute assessment administered to household members aged between 12 and 72. The content of the test was based on the national curriculum. Four versions of the test were used, depending on a respondent's age and education level. To improve the quality of the testing instruments, early versions of the test were piloted, some items were removed and others were altered. Item response modelling was used to combine the test into a single scale of mathematics literacy. This procedure included validity checks to ensure that the scale was internally reliable. Please refer to the NIDS Wave 1 User Document for further details about the variables derived from the numeracy assessment (http://www.nids.uct.ac.za/home/index.php?/Nids-Documentation/documents.html).

REFERENCES

Department of Education. 2005. "Grade 6 Intermediate Phase Systemic Evaluation Report." Department of Education, Pretoria.

Moloi, M. and J. Strauss. 2005. "The SACMEQ II Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education." SACMEQ, Harare.

Mullins, I.V.S., O. Martin, E. Gonzales, and S.J. Chrostowski. 2004. "TIMMS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades." TIMMS and PIRLS International Study Center, Boston.

UNESCO. 2008. "Education for All Global Monitoring Report." UNESCO, Paris.

Wright, B.D., & Masters, G.N. (1982). *Rating scale analysis.* Chicago: MESA Press.