

Trends in Student Performance Measurement

USAID Education Overview Course
for Education Officers
Bethesda, Maryland, May 14-25, 2007

Further questions, comments? Write lcrouch@rti.org or
for materials visit www.eddataglobal.org

Issues to be covered: what are the key trends?

- Key trends seem to be all over the place, often in (apparently) contradictory directions
- But only apparently
- Makes it harder to understand what is going on
- This presentation/discussion: help sort some of this out

NOTE: We will not get through this in 45 minutes.

Much of this will not be covered in the presentation but is given only as reference material!

Issues to be covered: what are the key trends?

1. Both more national and international (and regional)
2. Both towards more complexity and more simplicity
Is this a contradiction?
3. Both for teacher and system empowerment and support and for accountability
4. Both high-stakes and low-stakes (probably more the latter?)

Trend 1: National, international, regional trends

- More national assessment:
 - E.g., Latin America, 20-25 years ago, had hardly any (maybe Costa Rica)
 - Today nearly all countries have some
- Yet also more international:
 - “Big three”
 - TIMSS (Trends in International Math and Science)
 - PISA (Programme in International Student Assessment)
 - PIRLS (Progress in International Reading Literacy Study)
 - TIMSS: in 95, 8 developing countries; in 99, 16; in 2003 21
 - Some countries e.g. in Middle-East “adopting” TIMSS standards
 - PISA not quite as successful in expanding coverage
 - PIRLS only one round



ferrer assess in lat
am.pdf

Trend 1: National, international, regional trends

- And, more regional:
 - At least 3 “big” regional ones
 - SACMEQ (Southern and Eastern African Consortium in Measuring Educational Quality)
 - PASEC (Programme d’analyse des systèmes éducatifs de la CONFEMEN¹) for Francophone Africa
 - SERCE (Second Regional Comparative and Explanatory Study – form the Lab Latinoam Eval Calidad Educ)
- Growth of regional
 - SACMEQ went from 6 in SACME I 1, to 14 in SACMEQ II, now SACMEQ 3 in 15(?)
 - LLECE/SERCE 12 in 1997, 17 in 2006 (+1 Mexican state)



sacmeq.pdf

¹ Conférence des Ministres de l’Education des pays ayant le français en partage

Trend 1: National, international, regional trends

Web sites

- SACMEQ <http://www.sacmeq.org/>
- PIRLS <http://timss.bc.edu/pirls2001.html>
- PISA <http://nces.ed.gov/surveys/pisa/>
- TIMSS <http://timss.bc.edu/>
- LLECE / SERCE <http://llece.unesco.cl/documentosdigitales/>
- PASEC http://www.confemen.org/rubrique.php3?id_rubrique=3

Trend 1: National, international, regional trends

- Some benefits:
- Int, regional: force up quality of national
 - Dissemination of standards
- Regional: more trust, comfort
- Some regional now anchored to international (SACMEQ “links” items with PISA)
- Could anchor national to regional or international
- National have some serious weaknesses sometimes (e.g., Uganda: not reliable for comps over time)



cross national
studies.pdf

Trend 2: More complex and simpler

- More complex
 - Better standards: increased (or better measures of) reliability, validity, discrimination, comparability over time, richness of information, such as analysis of biases, and so on
 - IRT or latent trait analysis makes much of this possible (“Rasch”: a special type)
- Simpler
 - More directly meaningful to parents, teachers
 - “Classical” testing theory

- Summaries of basic issues in testing:



wb documentation
on testing.doc



basic concepts.pdf



More valid, less reliable



More reliable, less valid

Trend 2: More complex and simpler

- Examples
- More complex: most “standardized” tests, international tests
- Simpler examples
 - Many tools for monitoring student progress
 - Applied directly by teachers or supervisors
 - Early-Grade Reading examples follow
 - Reliability and validity based on simplicity of the task (discuss) more than on psychometric properties

Trend 2: More complex and simpler

- Examples
- More complex: most “standardized” tests, international tests
- Simpler examples
 - Many tools for monitoring student progress
 - Applied directly by teachers or supervisors
 - Early-Grade Reading examples follow
 - Reliability and validity based on simplicity of the task (discuss) more than on psychometric properties

Trend 2: More complex and simpler



zambia primary
reading program.pdf

- Zambia Primary Reading Programme

To measure: The ability to read simple texts such as letters, local language newspapers, books and messages. This ability is an important element in communication.

The reading test was given to one pupil at a time. The test administrator gave the pupil the first page of the reading test in a Zambian language and asked that the pupil should start reading and carry on reading until the pupil could no longer read. A pupil who coped adequately with the first page was given the second page of the reading test. There was a time limit of five minutes to read.

Same passage given across grades

Trend 2: More complex and simpler

- **Zambia Primary Reading Programme test**
 - single words: allocate one mark per word read correctly (1/2 a mark can be awarded for a good attempt)
 - sentences: allocate two marks per sentence read entirely correctly. If only a few words are read correctly from the sentence allocate ½ a mark, if at least half the sentence was correct allocate one mark and if there was only a slight mistake allocate one ½ marks.
 - Paragraphs: allocate eight marks for a paragraph read correctly. When allocating marks for paragraphs, use the same model as under b and c above, deducting one mark for every word read incorrectly and ½ a mark for slight mistakes.
 - Comprehension questions based on the last two long paragraphs in the test, should be read by the tester. Children should gain two marks for each full and correct answer, they can score one mark for a less full answer. We suggest that you read through these first yourself so that you know what type of answers to expect.

Mean performance in **Zambian languages for Grade I**

Province	1999	2002
Central	2.2	12.1
Copperrbelt	1.7	19.7
Eastern	1.6	20.9
Luapula	1.7	17.7
Lusaka	2.7	24.1
Northern	3.8	15.4
Northwestern	3.4	14.8
Southern	0.8	7.2
Western	0.4	7.8

Trend 2: More complex and simpler

- **USAID Early Grade Reading Assessment**
 - Pilot simple but rigorous test
 - Pilot a short form and a longer form
 - Short form: a) letter recognition, b) familiar words, c) nonword decoding, d) phoneme segmentation, e) passage reading for fluency and comprehension
 - Longer form helps validate shorter form
 - Assess “grade of breakthrough”
 - Assess reasonable expectations (“standards”) for key languages based on research itself
 - Compare within key languages but not “league tables” (?)



Early Grade
Reading Assess.doc

Trend 3: Both empowerment and accountability of teachers, schools

- **Empowerment: providing teachers and teacher training systems with feedback**
 - Requires detailed analysis, tight feedback loop
 - Cases: Uruguay, DDSP and IEP projects in SA
 - It is reason for simpler tests, may also require longer (but simpler) tests
 - Can be done on sample basis if patterns of failed items are common
 - Providing system with “factors” associated with learning (“production function” sorts of studies, feed more macro policy)
 - Key factors include: gender, social class, resources, etc.
 - Gender: where are boys falling behind relative to girls, where are girls falling behind???
 - Where is income more important than gender as explanatory factor?
 - How about language of teaching and learning?

“Empowerment” requires detailed analysis, item by item...

And may require lots of items in the test, to pinpoint conceptual problems teachers are leaving unresolved.

Of course it requires tight feedback loop from measurement to in-service training

Example of detailed analysis... feedback provided to teachers can be this detailed

Task	Sub-task	Problem	2000	2004	Gain
Counting & ordering	number line	identify no. 12 on line with 9 to 11 and 16 already labelled	29	56	+27
Counting & ordering	skip forward by 2: <100	34 36 38 ?	48	72	+24
Counting & ordering	skip backward by 10: <100	80 ? 60	38	62	+24
Counting & ordering	skip forward by 50: >100	250 300 ? ?	11	32	+21
Addition	>100, carrying, no context	$50 + 60 = ?$	19	37	+18
Counting & ordering	skip backward by 100: >100	570 470 370 ?	8	25	+17
Subtraction	>100, no carrying, no context	$115 - 15 = ?$	24	40	+16
Multiplication	no context	$10 \times 6 = ?$	43	59	+16
Addition	>100, carrying, no context	$240 + 60 = ?$	14	29	+15
Multiplication	no context	$2 \times 9 = ?$	46	61	+15

Workbooks and assessment resource banks provided to teachers, as a way of giving flesh to the concepts they are missing

Counting and Ordering

Test instrument

Count forward in 2s and fill in the missing number: 34 36 38 _____

ISP workbook

Count forwards in 2s: Fill in the missing numbers: 38 _____ 44 46 48

HSRC ARB

Count forwards in 3s and fill in the missing numbers: 18 21 24 _____

Addition

Test instrument

$34 + 8 =$

ISP workbook

$28 + 6 =$

HSRC ARB76 + 7 =

Subtraction

Test instrument

$28 - 7 =$

ISP workbook

$77 - 6 =$

HSRC ARB48 - 6 =

Multiplication

Test instrument

$10 \times 6 =$

ISP workbook

$6 \times 10 =$

HSRC ARB7 x 5 =

Word Sums

Test instrument

Nomsa has 8 bags of 10 oranges each. How many oranges are there altogether?

ISP workbook

6 piles of stones with 10 stones in each pile make _____ stones altogether.

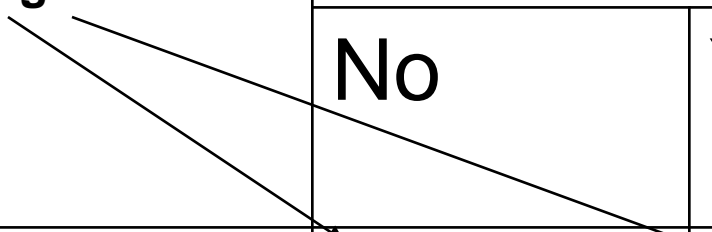
HSRC ARB1

worm eats 8 leaves each day. 15 worms eat _____ leaves each day.

Typical “associated” factors studies cover issues such as gender, income...

Help with pedagogy, also more macro policy issues

**For example, evidence on importance of home language instruction
(from SA's own national assessment)**

<p>African kids taught in their own home language score 100% better even in the DOMINANT language!</p> 		<p>Match Lang of Instruction and Home Language</p>	
		<p>No</p>	<p>Yes</p>
<p>Home Language</p>	<p>African</p>	<p>31%</p>	<p>61%</p>
	<p>White</p>	<p>46%</p>	<p>68%</p>

Or, gender issues

	% countries girls > boys	% countries boys > girls	% countries girls=boys
Grade 6 reading	40	20	40
15-yr old reading	98	0	2
Grade 6 math	14	43	43
Grade 8 math	20	20	60
15 yr old math	2	68	30

Pause for word on role of USAID in fostering more sophisticated national testing...

Example from DR...

EERC (Ed Eval Res Consortium):

- **Evaluate / understand USAID interventions**
- **Provide TA to DR on measurement**
- **Develop outcome indicators**

Uses state-of-art procedures and reporting on many interesting factors such as:

- **Gender (girls > boys)**
- **Private – public (private > public)**
- **General slowness: most of grade 5 has not mastered grade 3 curric**
- **Detailed error pattern analysis (see examples above too)**



**EERC presentation
at CIES 2007.ppt**



**EERC note CIES
2007.doc**

Trend 3: Both empowerment and accountability of teachers, schools

- **Empowerment: providing teachers and teacher training systems with feedback**
 - Requires detailed analysis, tight feedback loop
 - Cases: Uruguay, DDSP and IEP projects in SA
 - It is reason for simpler tests, may also require longer (but simpler tests)
 - Can be done on sample basis if patterns of failure are common
- **Accountability: providing parents, bureaucracy, with information on school (teacher?) performance**
 - Careful with perverse incentives
 - Requires universal testing
 - Often uses high-stakes universal student (filter) exams
- **Accountability without empowerment, and empowerment without accountability are probably both weak: need both**

Accountability systems may be simple but may be quite effective...

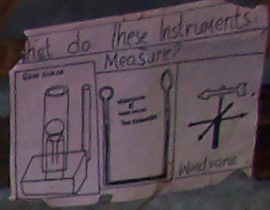


PLEASE DO NOT FEED THE ANIMALS
IN THE ZOO. FEEDING IS THE RESPONSIBILITY
OF THE VISITOR. THE ZOO IS NOT
RESPONSIBLE FOR ANY INJURY TO
THE ANIMALS OR VISITORS.

10000 = 100



$N = \frac{P \times R \times T}{100}$
 $P = \frac{N \times 100}{R \times T}$
 $R = \frac{N \times 100}{P \times T}$
 $T = \frac{N \times 100}{P \times R}$
 $A = I + P$



ENROLLMENT

CLASS	BOYS	GIRLS	TOTAL
ONE	41	30	71
TWO	30	42	72
THREE	17	24	41
FOUR	22	25	47
FIVE	14	16	30
SIX	17	19	36
SEVEN	14	19	33
EIGHT	13	6	19
TOTAL	168	177	345

PUPILS TO
PREPARE FOR
MID-TERM
EXAMS

SCHOOL BASED TEST RESULTS

END OF TERM II 2005

CLASS	ENG	KISW	MATH	SCIE	CHRY/5/SR	TOTAL	POSITION
1A	27.05	24.00	32.02	16.2	38.5	165.80	
1B	26.98	25.01	33.00	18.5	39.6	164.04	
2A	29.3	25.0	36.3	19.0	40.23	169.60	
2B	28.0	23.02	30.3	19.15	40.55	161.03	
3	25.92	25.15	17.48	15.97	32.5	141.46	
4	40.98	46.95	40.78	36.78	34.39	199.88	
5	43.58	49.24	41.34	41.07	46.10	221.90	
6	42.46	57.26	42.06	41.74	44.09	227.60	
7	43.44	52.94	40.76	48.01	52.72	228.86	
8	50.42	56.05	49.47	52.95	57.79	267.12	

YEAR KCPE RESULTS

YEAR	ENG	KISW	MATH	SCIE	CHRY/5/SR	TOTAL
2001	50.18	60.45	55.91	51.46	56.72	277.71
2002	48.92	53.71	44.14	44.57	48.21	239.55
2003	47.46	61.75	51.01	49.88	45.21	255.31
2004	46.52	49.30	41.25	44.74	47.00	228.81

STAFF
MATCHMA
F.P

K.C.P.E MEAN SCORE

2002	166.27
2003	195.6
2004	189.60
2005	252.56
2006	
2007	
2008	

SCHOOL ENROLM

CLASS	1	2
BOYS	119	92
GIRLS	64	55
TOTAL	183	147

SCHOOL FACILITIES

CLASSROOMS	20
------------	----

2005 F.P.E. DISBURSMENT OF FUNDS

GENERAL PURPOSE ACCOUNT

SIMBA ACCOUNT

Vote Head	1st		Debit
	2 nd	3 rd	
SS.W	58605	34000	46005
RMI	255250	246509	8741
ACT.	22077	9400	12677
Q.A.	15831	12312	3519
LT&T	11946	8200	3746
EWC	6716	400	6317
TEL/POST.	11220	4200	7020
CONT.	3159	2194	965
TOTAL	384805	337245	47590

Vote head	
Text books	72,600
Ex. books	36,300
Supplementary	11,616
Pencils	3630
Duster/chalk	1452
Charts/w. Maps	1452
Total	127,050
2 nd Dis.	53,800

USED
BANK

Trend 4: High stakes or low stakes?

- **In my view, I don't see much increase (if any) in high-stakes testing**
- **Not for students anyway**
- **Few countries increasing the use of “filter” exams**
- **Some (well, one?) create stakes for teachers (e.g., merit pay) using exams**