# Possibilities for comparing existing international, regional and national surveys

*Ray Adams, November 2014*

Australian Council *for* Educational Research

# An apology on behalf of psychometricians (or perhaps to them)

**Perfect is the enemy of good**

striving to better, oft we mar what's well

Give them the third best to go on with; the second best comes too late, the best never comes

If you never miss a plane, you're spending too much time at the airport

*http://en.wikipedia.org/wiki/Perfect_is_the_enemy_of_good*

# Psychometricians' Concerns

- Technically robust and globally comparable indicators
  - The construct to be measured must be valid and reliable across all countries
  - ideally measured by a similar question or item
- Things we can measure are given prominence and value over and above things we cannot measure

# But there are problems that need to be solved

## Muscat Agreement

By 2030, all girls and boys complete free and compulsory quality basic education of at least 9 years and achieve relevant learning outcomes, with particular attention to gender equality and the most marginalized.

By 2030, all youth and at least x% of adults reach a proficiency level in literacy and numeracy sufficient to fully participate in society, with particular attention to girls and women and the most marginalized.

## Open Working Group

by 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes

by 2030 ensure that all youth and at least x% of adults, both men and women, achieve literacy and numeracy

# The Way Forward

- Develop learning metrics to form the backbone of outcome indicator régimes
  - Goals, targets, benchmarks and indicators only make sense if they relate to a supporting and underpinning framework (metric)
- We must take a bottom up approach
  - build a metric from existing metrics
  - expand existing metrics
  - Quality within-country measurement practices should not be compromised
- Measure of success is fitness for purpose…not perfection!

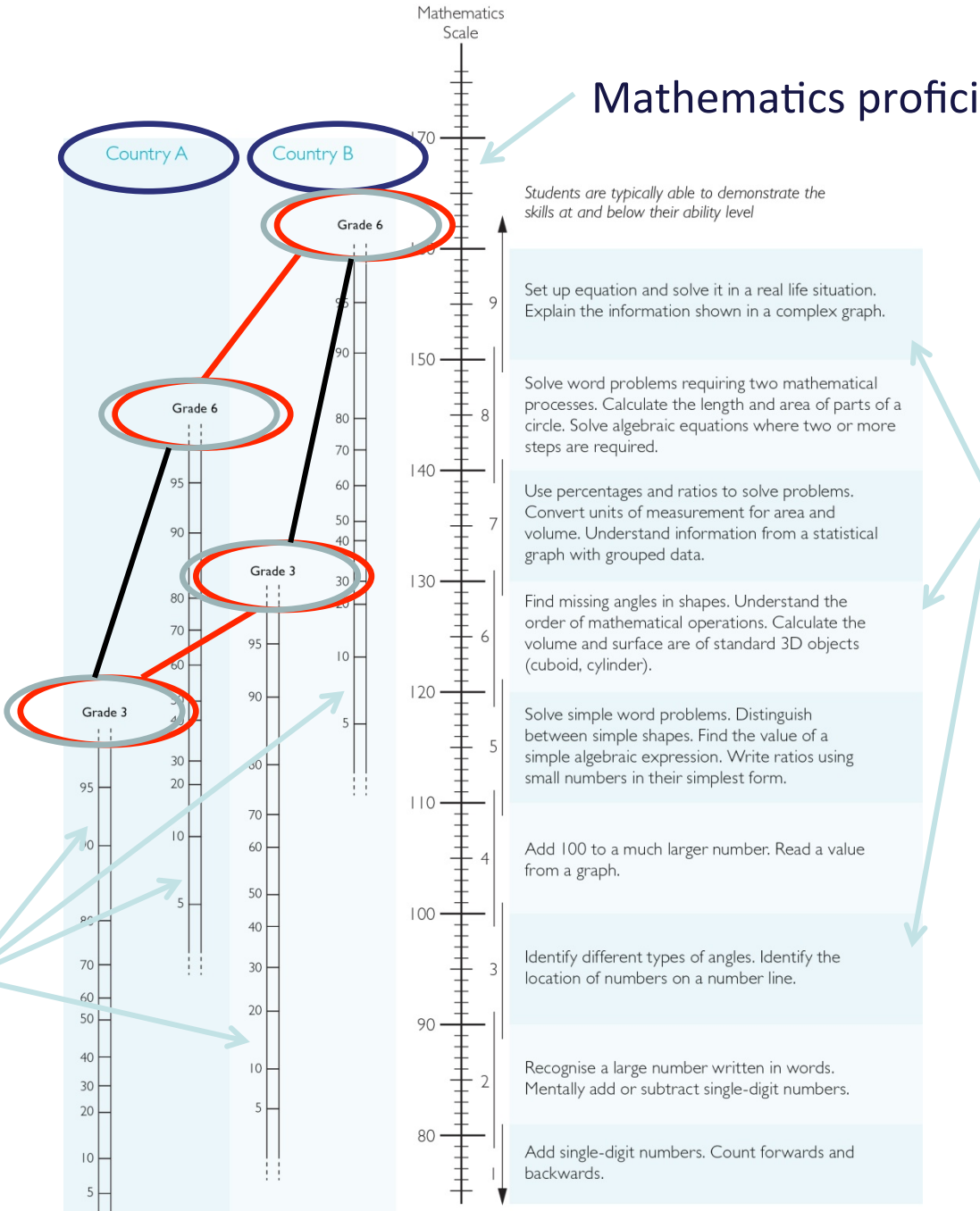
# What is a learning metric?

- A metric is a dimension of educational progress
  - depicted as a line showing ‘more’ and ‘less’
- Locations along the described numerically are proficiency scores
- Locations along the scale described substantively are proficiency descriptions
- The metric is not bound by objects of measurement e.g. independent of age/grade

# Benchmarks and indicators

- Benchmarks are points on the scale, that include a statement about typical, acceptable or desirable performance
  - "score X is an acceptable minimum performance level for students at end of primary school"
- Indicators are quantitative expressions that use the metric to provide information in relation to targets and goals.
  - "the proportion of students who can meet specified benchmark"

Mathematics Report for Country A and Country B 2012




Mathematics proficiency scale

Growth comparisons

Cross-country comparisons

Percentile distributions for tested populations

Descriptions of proficiency levels

# Phase 1: Draft Metrics
## (to be completed end 2014)

- Analyse the substance of existing assessments, including **frameworks**, **data** and **items** to see how well they can be aligned to a common metric
- No new student data – a conceptual approximate alignment that provides a **draft metric** and mapping of some tests to it
- Supported by a pairwise comparison study

# Phase II: Metric validation (30 months)

- Collection of new data to empirically validate the draft metrics from phase 1
- Would require utilisation of instruments used in phase 1 in equating studies that need to be designed
- Product would be
  - Validated metrics
  - Comparable information for countries that have participated in contributing studies

# Phase III: Implementation (ongoing)

- Work with countries ... in depth localised system strengthening ... to assist with developing **sustainable, high quality** learning assessment systems that inform policy development
- Links to the metrics built into their systems, using a variety of suitable and locally acceptable methodologies

# Phase III: Link Methodologies

- Build link items into their collections
- Participate in a regional or international learning assessment
- Undertake an equating between local assessment and previously equated assessment
- Need for quality assurance for acceptance of mapping to global metric

# Progress To Date – Phase I

- Funded by DFAT (Australian Government) and ACER
- Established a team of curriculum and test development experts for reading and mathematics
- Collected items and frameworks form many assessment programs -- PISA, TIMSS, SISTA (Solomon Islands), PILNA (Pacific Islands), EGMA, EGRA, LLANS, ASER (India), Uwezo (Uganda), OLE (Northern Territory online assessment), Zimbabwe and Afghanistan.
- Prepared descriptions of the typical conceptual and skill development sequence in mathematics and reading, from early primary through to early/middle secondary levels.
- Analysed item demand for the items and develop a description of the main elements of cognitive demand for each item and mapped those to the developmental sequences.
- Use data from past administrations of those items to empirically determine the relative difficulty of items within each assessment program.
- Conducted a pairwise comparison study to check relative item difficulties across different assessments.
- And now, the reading and mathematics teams are building the draft learning metrics by laying out items and their demand descriptions along a line (for each of reading, and mathematics), and will shortly develop draft descriptions of regions of the scale.

# Progress To Date – Phase II and beyond

- An LMP Secretariat formed as a collaboration between UIS and ACER
- A steering group and a technical reference group formed to:
  - Provide technical guidance and quality control
  - Advise on strategic objectives
  - Provide political support
  - Facilitate cross-country exchange and partnerships
- Detailed phase II plan in preparation
  - Addressing issues such as: how many countries and their range, number of items, sampling and scaling processes, translation, quality assurance, capacity building plans
- Strong need to engage with regional assessments (SACMEQ, PASEC, LLECE, SEA-PLM) and IEA and OECD

# Issues and Risks -- 1

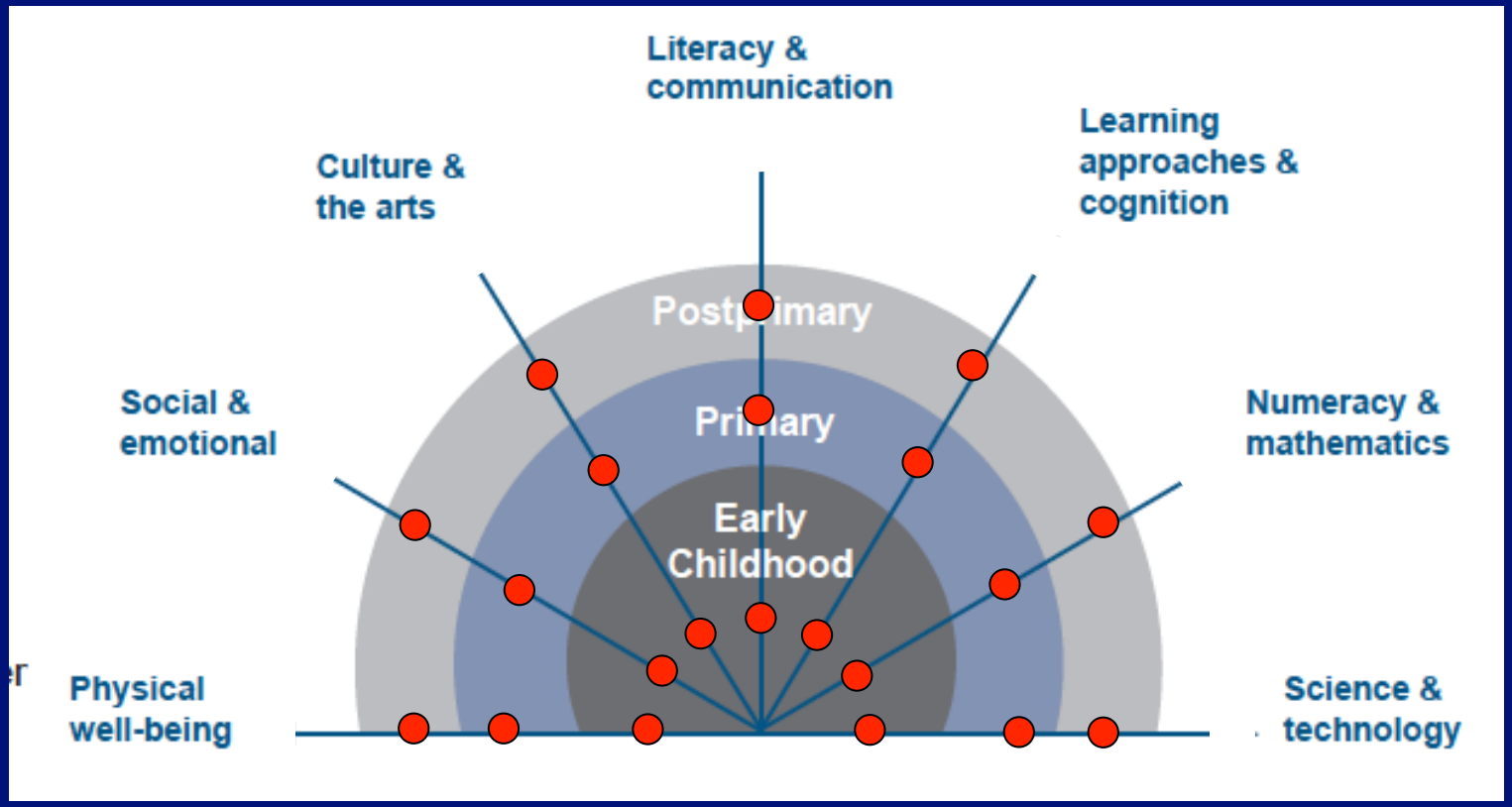
- Credibility
  - Agreement on what should be measured
    - the learning metric
  - Acceptance of comparability
    - technical rigour
    - methodological/method influences
  - Agreement on standards
    - successful completion?
    - age appropriate?

*Data are only as good as the system (or process) that collects them. Analysis is only as good as the data on which it is based and the skills and experience of the analyst.*

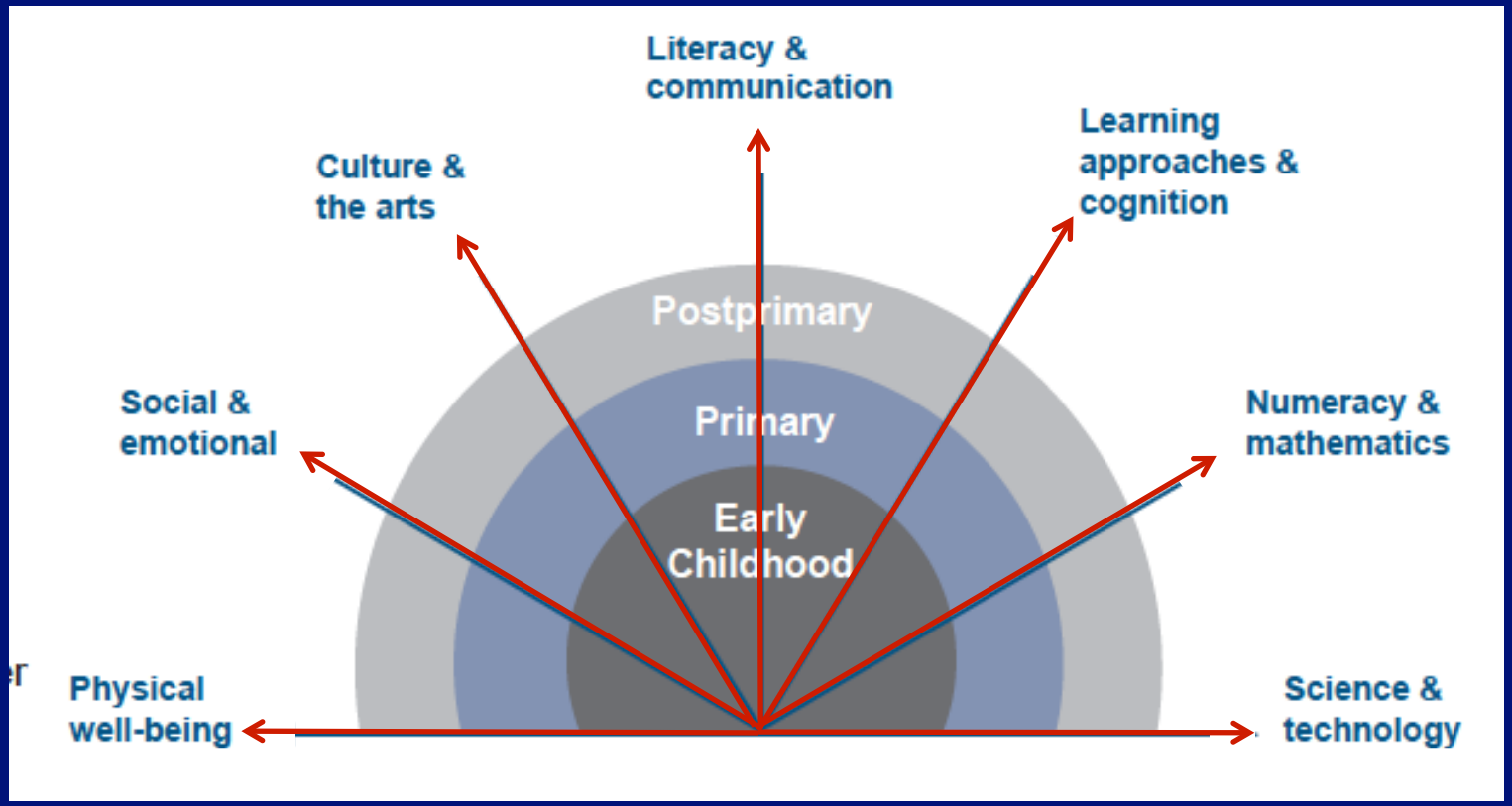
# Issues and risks -- 2

- Accessing data for phase 2
  - Engaging regional assessment
  - Engaging national assessments
- Willingness of countries to locate themselves on the metric
  - Political
  - Concerns about quality/validity
- Financial support
  - International and national
- Challenges/feasibility of quality assurance

# Three Concluding Points
# 1 – It is all about growth



# Three Concluding Points
# 1 – It is all about growth



# Three Concluding Points
# 2 – Assess where students are, not where you would like them to be

http://teacher.acer.edu.au/geoff-masters/article/achieving-high-standards-by-starting-from-current-performance

# Three Concluding Points
# 3 – Assess what students can do with their knowledge

All curriculum has a purpose

Providing learners with the capacity to deal with authentic problems

Empowering beyond the classroom

To be life long learners

*http://www.acer.edu.au/files/The_literacy_idea_2014.pdf*