GEM THE CENTRE FOR GLOBAL EDUCATION MONITORING

ASSESSMENT GEMS SERIES

No. 9

December 2015

Monitoring reading progress: Towards a global approach

Ross Turner, Juliette Mendelovits, Prue Anderson, Ray Adams

Australian Council *for* Educational Research

ACER

ISSN 2203-9406 (Online)

This paper explores some of the opportunities and challenges confronted by researchers and service agencies seeking to support the international community to monitor progress against the United Nations Sustainable Development Goals. The paper's main focus is on measuring children's progress in learning to read.

The paper discusses some of the characteristics of an effective approach to measuring outcomes in a globally consistent way; describes progress in beginning to implement an approach to achieve this; and mentions some of the ways current assessment and measurement activities might relate to this approach.

Effective measurement of global reading outcomes

The ability to provide meaningful measures of reading growth in students across the globe will be a critical element of driving progress towards the Sustainable Development Goals over the coming 15 years.

This means taking seriously the measurement challenges in relation to the Sustainable Development Goals (SDGs), in particular goals 4.1 and 4.2:

> *4.1: By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes*
>
> *4.2: By 2030, ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education*[1]

We need to be ready to provide meaningful comparative data from early childhood through primary education to secondary level. The tools that will be needed by the international community to monitor progress against the SDGs across the required span of the learning spectrum currently do not exist. Similarly, there is no clear consensus on what 'effective learning outcomes' might be. This is our challenge.

Effective measurement in support of the SDG process begins with the development of a metric for each learning area of interest. This means developing a means of describing and quantifying progress along a trajectory of learning growth in that area. It means developing a useful set of measures of that growth.

Learning outcomes lie along a continuum. Progression along that continuum occurs in a somewhat lumpy way, but sees individuals learning a particular language gradually progressing through a more or less consistent set of steps, in a more or less consistent order, but with each step negotiated at different speeds for different individuals. The differences between individuals' progression are affected by factors such as readiness (which in turn is affected by the nature of previous experiences), current resource availability and application, health and other personal factors, occurrence of learning opportunities, effectiveness of teaching strategies, and so on.

An essential methodological step underpinning effective measurement of reading ability is to describe that learning progression, and to recognise differences that may exist in this progression across languages of differing kinds. Different kinds of reading tasks are designed, and student accomplishments in relation to those reading activities are measured or described. Reading tasks must be designed to provide indicators of the reading ability of individual students at every stage of their reading growth in order to adequately measure that growth across the full range of students of interest.

The measures and the metric should be capable of helping to identify small learning steps, growth over a period of time, comparisons of progress made by different groups, as well as progress against defined achievement benchmarks. The measures should provide information that is meaningful and consistent across a wide range of ages, locations and language types.

Moreover, the metric should provide information that indicates where learning interventions should be targeted.

Well-constructed metrics for reading and mathematics may have other uses in some settings. For example, they might be useful as part of a review of teacher preparation programs or other teacher professional development courses, to ensure that relevant pedagogical knowledge is

1 For the full list of SDGs and targets, see the United Nations Sustainable Development Knowledge Platform at https://sustainabledevelopment.un.org/focussdgs.html

covered. In addition, they may also provide part of a framework in the review of curriculum documents.

Defining and measuring reading growth

ACER's Centre for Global Education Monitoring[2] (ACER-GEM) has been working to articulate a number of principles and lessons learned over recent years of involvement in a wide range of assessment programs in different countries around the world that frame our thinking and work about the development of a global metric for reading. These principles align well with the thinking that underpins the SDGs.

There are three central elements to the approach adopted by ACER-GEM when developing assessments of student outcomes. The first is the view that student assessment should be targeted *to identify where students are* in their progress along well-defined learning metrics. This approach contrasts with assessments that are oriented to gauging whether students have learnt the content of the curriculum considered appropriate for their age or year level. At the core of the development of strategies for improving learning is an understanding of where students are in their educational progress, not a statement of how far away they are from where you would like them to be.

The second element of the ACER-GEM approach is the view that an effective curriculum develops student skills, knowledge and understandings for the purpose of taking those skills, knowledge and understandings beyond the classroom and into all theatres of daily activity, beyond schooling, to work and other areas of life. Such an education is said to have a *literacy* orientation.

The third element is the view that growth along well-defined learning metrics for all learners is the goal of all education systems and practitioners. We see *growth for all*, regardless of their starting point or their background, as the core of an equitable approach to education. Consequently we see the gathering of data about growth over time – for individual students and cohorts of students – as a core element of monitoring educational outcomes.

2 ACER's Centre for Global Education Monitoring is a strategic initiative of the Australian Council for Educational Research, and is supported by funding from both ACER and the Australian Aid Program of the Australian government's Department of Foreign Affairs and Trade.

Steps for successful measurement

To be most useful, those measurements need to be comparable across different measurement contexts – different places, different measurement occasions. Locations on a learning metric should mean the same thing whether they are derived from measures of progress generated in Afghanistan, Angola or Australia; learning progress against the SDG goals should be expressed in units that have the same interpretation everywhere. In addition, they should be meaningful – they should help the user (policymakers, and ultimately teachers) see what needs to come next in the learning venture.

A potentially workable approach to this task builds on a methodology seen in various places around the world. For example, the National Assessment of Educational Progress in the United States has used a form of learning metrics in its reporting for many years, as have the studies by the International Association for the Evaluation of Educational Achievement (IEA) (Trends in International Mathematics and Science Study and Progress in International Reading Literacy Study), and the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA).

The first step is to describe the domain of interest. Let us assume that we are talking about reading, and we have laid out what we understand to be the key elements of reading proficiency, have described the major steps and markers in the process of development of reading proficiency, and have devised an operational definition that shows how indicators of reading growth can be generated.

Reading experts translate the framework and the indicators identified into a set of test items that are capable of showing different levels of reading proficiency across every identifiable phase of reading growth. These can be developed by item writers, or acquired from existing pools of items.

Operational work follows, whereby items from a large pool are administered to a large number of learners, preferably from the widest possible range of different contexts. The purpose of this work is to generate empirical data to inform building of the metric.

Technical work is then needed to calibrate the learners and the items and place them on a single scale (using methods from Item Response Theory).

The scale is divided into bands or levels, and the descriptions of items within each level are used to build a generalised description of the skills and knowledge located at each band. This allows outcomes to be more easily interpreted.

Such a process – using a sufficiently wide range of test items, and targeting a sufficiently wide range of aspects of developing reading proficiency – can be used to generate a global learning metric. This metric would be a tool that could help users to develop a common understanding of where students are in their learning growth. What is special about a global metric is that it offers information that is comparable across different languages, different stages of learning growth and different country contexts.

We need such a tool in order to meet the emerging measurement needs of the development sector around the world. ACER has commenced this task, by working with a range of different kinds of existing assessments – community-based, local, national, regional and international.

Preliminary results

ACER-GEM has approached the building of a reading metric by recognising that reading comprehension (which is the main objective of learning to read) starts before students can read. The first steps in reading growth are indicated through oral expression of precursor skills, and through response to oral or other non-verbal prompts.

In measuring reading proficiency, it is essential to recognise that reading comprehension:

- starts with recognising sounds, and their association with objects, events and concepts experienced by the individual
- continues through the development of an oral vocabulary, and gradual mastery of the structure of language as experienced orally
- progresses through a recognition of symbols and signs that have been invented by humans to represent sounds, objects and events
- develops as sounds and symbols are grouped together to form progressively larger and more complex language elements (words, phrases, sentences, and so on).

Increasingly advanced reading comprehension skills are indicated both by a growing vocabulary and by the individual's ability to make sense of increasingly complex expressions of language in an increasingly varied range of contexts.

However, different languages also present a range of challenges in the measurement of reading progress. Languages differ in the nature of the relationship between sounds and their associated symbols.

It seems clear that an effective approach to measuring progress in learning to read must start with oral language. It is also clear that attention must be paid to the granularity of any measurement scale that is used to monitor reading growth in order to accommodate differences in linguistic patterns of growth among different languages.

The ACER-GEM team has compiled more than 500 reading items (and a similar number for mathematics) from assessments conducted in Australia, Europe, the Pacific Islands, Africa, Southern, Central and Eastern Asia, and North and South America. Some of those assessments are able to provide good measures of the relative difficulty of items they use, but what we do not have is comparative information that allows us to locate the data from one assessment program on the same scale as data from another.

A preliminary means of generating such data is to use a long-standing methodology that involves trained expert raters making pairwise comparative judgments about the relative difficulty of the items. This has been done, using some 23000 pairwise comparisons, in order to get a first approximation of the location of all of these items on a single scale.

We have used the approach to building learning metrics, described earlier in this paper, to construct an initial draft metric spanning all of these items.

The pairwise comparison suggested that the OECD's Programme for International Student Assessment (PISA) had the most difficult items, and the Research Triangle Institute's Early Grade Reading Assessment (EGRA) had the least difficult – though both covered a wide range of difficulty. The Australian Longitudinal Literacy

and Numeracy Survey overlaps in difficulty with EGRA. The Annual Status of Education Report and its derivative assessments span a small range of difficulty at the lower end of the spectrum. ACER's Monitoring Trends in Educational Growth assessment conducted in Afghanistan for Year 6 students falls between PISA and the early grade assessments.

Our work suggests that the development of a global metric for reading needs to take account both of the unconstrained skills (those that continue to develop: reading comprehension and vocabulary development), and of the constrained skills (those that are learned relatively quickly and mastered entirely). Concepts of print, phonological awareness and reading fluency are included in this latter category. Indeed it is important to recognise that the steps and processes involved with learning the constrained skills that underpin so many national and international early reading assessments may be only loosely related to the steps and process involved in developing reading comprehension. Assessments must take full account of these two aspects of reading growth.

Our working hypothesis is that unconstrained skills are universal, hence a single comprehensive learning growth scale can be described. By contrast, the constrained skills are language-specific.

It has become evident that different languages – depending on their complexity, regularity and orthographic transparency – will take longer or shorter for children to master as part of their developing text comprehension. For example, children take a relatively long time to become fluent decoders in English, because it has a substantial amount of irregularity. Our first attempts to research the development of Japanese reading (Hiragana and Katakana) suggest that, because it has quite regular symbol-sound relationships, it takes less time for children to master the constrained skills; and they reach a ceiling earlier than English constrained skills in relation to the development of text comprehension. Spanish, too, has a very transparent and regular orthography, so decoding is mastered early. Arabic, although it has regular orthography, has the complication of tashkeel (dots), which are excluded in adult texts, so mastery of decoding in Arabic typically progresses more slowly in relation to text comprehension than in some other languages.

Another particular set of challenges emerge in relation to the earliest stages of language learning. It seems unhelpful to have measurement tools that tell us the majority of early learners in developing countries have minimal decoding skills, and therefore according to current assumptions, have zero reading comprehension skills.

ACER's experience shows that it is possible to measure very relevant precursor skills, which are highly correlated with later reading proficiency. This can be done using a range of semantically simpler tasks than are commonly used at present, for example matching a written word to a picture, or (for oral comprehension) asking a child to say what a picture shows. Provision of sound files has been successfully used with some young readers to direct their attention to particular stimulus and to deliver questions as part of a computer-based assessment methodology. More advanced reading comprehension tasks involve varying the length and complexity of written stimulus. To be most effective they avoid confounding memory with comprehension. We seek to better discriminate among very early learners who are near the beginning of their reading growth, while nevertheless at different stages.

Our work to develop metrics for reading and mathematics has only just begun. We have developed some initial draft material, and these products need to be validated empirically in a range of countries. This should include validation activities using a variety of languages, in order to confirm the relationships among items across different assessment programs, and in the case of reading, to advance our thinking about the variations in constrained skills and how these relate to the development of text comprehension.

Once developed, further empirical work will be needed to build mechanisms for aligning existing national, regional or international assessments with the global metric. The intention is to jointly develop a methodology that can be used by the international community to monitor progress against the education-related SDGs. For this work to be most effective, it must be done in partnership with and in support of the efforts of the various bodies having an interest in the global development sector.

Work to meet the 2030 United Nations Sustainable Development Goals should focus on the development of measurement tools that:

- capture learning growth (at least in reading and mathematics, but other areas may also be needed) with a focus on where students are in their learning rather than where they 'should be'
- span a broad range of learning growth stages, from early childhood through to secondary school
- capitalise on current assessment activities
- provide information that is sufficiently comparable for the required purposes, across a broad range of contexts
- describe what students can do, and what they need to do next in order for their learning to continue to develop
- take into account differences across languages.

The ACER Global Education Monitoring Centre supports the monitoring of educational outcomes worldwide, holding the view that the systematic and strategic collection of data on educational outcomes, and factors related to those outcomes, can inform policy aimed at improving educational progress for all learners.

www.acer.edu.au/gem