GEM THE CENTRE FOR GLOBAL EDUCATION MONITORING

ASSESSMENT GEMS SERIES

No. 4

June 2014

Described proficiency scales and learning metrics

Ross Turner

Introduction

We see the process of educational measurement as one of defining dimensions of educational progression and locating learners on those dimensions. Such dimensions are variously referred to as proficiency scales or learning metrics.

When we measure something using a scale we can express how much of a given attribute is possessed by that something. We use a scale of centimetres or inches to describe the length of an object. We use a scale of hours and minutes to describe an amount of time. This is also the case for educational measurement: when we measure proficiency in a subject area we use a proficiency scale (or learning metric) to describe the extent to which the learner possesses the skills, knowledge and understanding that comprise the area.

A learner who is further along the learning metric has greater proficiency in a subject area than a student at a lower point of the learning metric.

Our approach to this area sees learning metrics as comprising two main elements: measures of proficiency located along a scale, and proficiency descriptions associated with locations on the scale.

A learning metric details and describes the different levels of proficiency shown by learners in a particular subject area or 'domain'. The metric describes what learners know, understand and can do at different stages of their development. A key concept underpinning learning metrics is that learning involves building and developing knowledge, skills and understanding in that area of learning. A learning metric is a basic tool that is used to report progress in a learning assessment (Masters and Forster 1996).

A learning metric is based on the idea that learning is something that builds over time and that is continuously progressing. It assumes that achievement at a given level of proficiency incorporates the knowledge, skills and understanding described in all of the levels below it.

The metric is depicted as a line with numerical gradations that quantify how much of the measured variable (for example reading ability) is present. Locations along this metric can be described by numerical scores or substantively (that is, in terms of student skills, understanding and competencies).

ACER

Australian Council *for* Educational Research

ISSN 2203-9406 (Online)

When the locations are described numerically, they are referred to as *proficiency scores*, and they quantify different performance standards for the metric. For example, a score of 115 is a proficiency score. When locations are described substantively, they are referred to as *proficiency descriptions*. For example, in the case of mathematics, a proficiency description might be something like, *students with a score of 115* (say) *on the scale can solve simple word problems, distinguish between simple shapes, find the value of a simple algebraic expression and write ratios using numbers in their simplest form.*

In reality it is not practical to develop a proficiency description for each proficiency score on a numerical scale, so proficiency descriptions are usually developed to cover particular segments of the scale. These segments are called *levels*. The proficiency description for a particular level can then be understood as describing the skills and proficiencies of students who attained proficiency scores that are within that particular segment of the scale (and those students would also have the proficiencies described in all lower segments). For example, again in the case of mathematics, *students at level 5 can solve simple word problems, distinguish between simple shapes…*

We may choose to set a location on the scale as a *benchmark*, which is a point on the scale against which we would like to make comparisons. For example, we might say that the score of 115 (the proficiency score described above) is a benchmark for *acceptable performance after the completion of primary schooling.*

An *indicator*, at least in this context, is a quantitative expression that is used to describe the quality, the effectiveness, the equity or the trends of a particular aspect of the education system. It does so through mathematical statements concerning metrics, proficiency scores and benchmarks. For example, *the proportion of students that have achieved a score of at least 115 in mathematics* is an indicator. Further, given the proficiency description of this score, an equivalent indicator is: *the proportion of students that can solve simple word problems, distinguish between simple shapes, find the value of a simple algebraic expression and write ratios using numbers in their simplest form.*

An example of a learning metric of numeracy/mathematics is shown in Figure 1, which is based on a metric developed by ACER for use in an Australian project. The central elements of the learning metric are the numerical scale, which in this case runs from below 80 to 170 vertically up the page, and the descriptions of the nine levels or segments of the scale in meaningful substantive terms.

The boundaries of the described levels are arbitrary. They can be constructed to make the levels discrete to imply a clean step from one level to the next, or they can allow for overlap to emphasise that moving from one level to the next might entail displaying some of the characteristics of neighbouring levels.

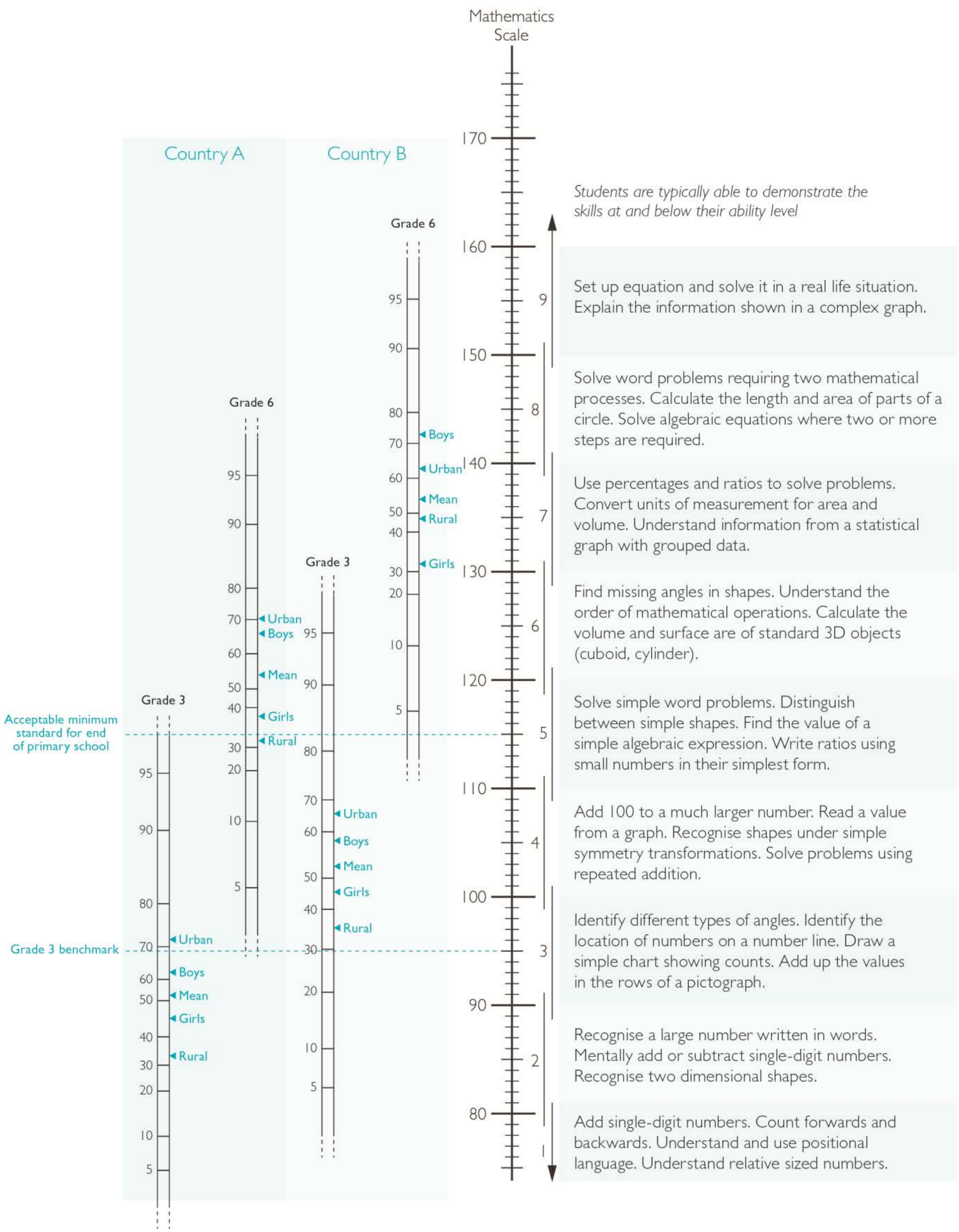
A described proficiency scale makes explicit what growth in an area of learning means. While a result on a numerical scale gives a quantitative report on the achievement of a student or a group of students, a described proficiency scale supports the interpretation of the score, in terms of the knowledge, skills and understanding typically associated with the score. In other words, it puts into words what the score means.

The various locations on this metric (or scale) are proficiency scores. Given agreement on the metric, assessment tools can be developed and locations on the scale can be chosen as benchmarks.

In Figure 1, the learning outcomes of two countries at Grade 3 and Grade 6 are reported in relation to the learning metric. For each grade for each country, a range of indicators is shown: the distribution of performance; the mean proficiency scores for all children; and the mean proficiency scores for girls, boys, urban children and rural children. A range of other indicators could also be highlighted – growth over years, differences between subgroups and so on.

Matching the mean proficiency scores of the different groups to the proficiency descriptions of the levels gives an understanding of the skills and abilities of these groups. The mean proficiency scores can also be compared to the acceptable Grade 3 and end of primary school benchmarks that are shown.

Figure 1: Example learning metric for mathematics



What are the benefits of this approach to learning metrics?

Both the development and use of learning metrics bring a number of benefits.

An education system can use a learning metric to better understand what progress in learning looks like in a particular learning area, and to understand at system or sub-group level the progress in learning that has been made. Teachers can use this information to help plan their teaching, by seeing what their students have already learned, and what they need to learn next. Parents can use this information to confirm what their child has achieved, and to anticipate what he or she will be working towards. Learners themselves can gain a better understanding of the learning they are experiencing.

Learning metrics provide a mechanism for expressing and sharing learning goals. For example, the process of setting and monitoring learning goals must have at its core a set of agreed learning metrics so that indicators that include terms such as *foundation skills* and *acceptable* (in terms of proficiency) can be used with the knowledge that they carry a shared and accepted meaning.

How is a learning metric created?

Creating an assessment involves an initial understanding and intentions about what improvement in the learning area looks like. At the core of the development of a learning metric is both an expectation of how proficiency develops and a view as to what would serve as evidence of that proficiency development. Learning metrics are then validated through analysis of actual learners' performance on sets of assessment tasks. In other words, learning metrics are based on conceptual understanding of a domain and educational intention, and then refined and supported by evidence.

The first phase in creating a learning metric is describing the extent of the domain and how intended growth in the domain would be evidenced.

The second phase requires the development of tasks that provide an opportunity to collect evidence concerning student proficiency in the domain of interest. To do this items are constructed that specifically tap aspects of the domain at various levels.

The third phase of development involves large numbers of learners attempting each assessment task. Following this, the psychometric analysis of learners' responses to the items allows the tasks to be located on a scale according to their level of difficulty.

Phase three can yield unexpected results, because in practice some tasks may not behave as intended. For example, a particular task may not turn out to be assessing the same broadly conceived subject area as the other tasks or may prove more or less difficult than expected from the preliminary estimation of its demands. The tasks that produce these unexpected results must be carefully scrutinised.

To finalise the learning metric, the tasks are arranged in order of difficulty and their descriptions are reviewed to identify the common features of groups of tasks with similar levels of difficulty. The task-level descriptions along with a conceptual understanding of the domain are then used to formulate summary descriptions of the kinds of proficiencies observed in each group of items, and described proficiency levels can then be defined on the scale.

Often, a selection of released items is used to illustrate the scale and its levels.

A simplified example of how a set of item descriptions is converted to a learning metric is provided in Figure 2. The set of item descriptions is from the thematic report on reading, *Reading for Change* (OECD 2002); the described level (Level 3) is reproduced in Volume I of the international report on PISA 2009 (OECD 2010).

Figure 2: Development of a level description for PISA Level 3 reading

542	INFER AN ANALOGICAL RELATIONSHIP between two phenomena discussed in an open LETTER.	Tasks at this level require the reader to locate, and in some cases recognise the relationship between, several pieces of information that must meet multiple conditions. Interpretive tasks at this level require the reader to integrate several parts of a text in order to identify a main idea, understand a relationship or construe the meaning of a word or phrase. They need to take into account many features in comparing, contrasting or categorising. Often the required information is not prominent or there are other text obstacles, such as ideas that are contrary to expectation or negatively worded. Reflective tasks at this level may require connections, comparisons and explanations, or they may require the reader to evaluate a feature of the text. Some reflective tasks require readers to demonstrate a fine understanding of the text in relation to familiar, everyday knowledge. Other tasks does not require detailed text comprehension but require the reader to draw on less common knowledge.	552
540	IDENTIFY the implied starting date of a GRAPH.		
539	CONSTRUE THE MEANING of short quotations from a LONG NARRATIVE in relation to atmosphere or immediate situation. (Score 1)		
537	CONNECT evidence from LONG NARRATIVE to personal concepts in order to justify opposing points of view. (Score 2)		
529	EXPLAIN a character's motivation by linking events in a LONG NARRATIVE.		
508	INFER THE RELATIONSHIP between TWO GRAPHIC DISPLAYS with different conventions.		
486	EVALUATE the suitability of a TREE DIAGRAM for particular purposes.		
485	LOCATE numerical information in a TREE DIAGRAM.		
480	CONNECT evidence from LONG NARRATIVE to personal concepts in order to justify a single point of view. (Score 1)		480

References

Masters, G.N. & M. Forster (1996). Developmental Assessment. Camberwell: Australian Council for Educational Research.

OECD (2002). Reading for change: Performance and Engagement across Countries: Results from PISA 2000. Paris: OECD.

OECD (2010). PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I). Paris: OECD.

OECD (2012). PISA 2009 Technical Report. Paris: OECD.

The ACER Centre for Global Education Monitoring 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.