But can we measure it?

It’s often asserted that some things can’t be measured. But how true is this? And if we can’t measure something, should we stop pretending we can teach or develop it?

Can creativity be measured? What about resilience? Can we measure problem solving? What about ethical reasoning?

We take for granted the measurement of variables such as length, weight and temperature, but it is worth remembering that these variables and their measurement were human inventions. Albert Einstein made this point in the Evolution of Physics (https://www.google.com.au/imgres?imgurl=http://t2.gstatic.com/images%3Fq%3Dtbhn:ANd9GcQSsWWI9jKvgG3ASm6vgDtHe81sWZrURr-COPY0IVAKVAwdw&imgrefurl=https://books.google.com/books/about/Evolution_of_Physics.html%3Fid%3DIWEmNBaHCJMC%26source%3Dkpc_cover&h=2255&w=1400&tbnid=2r10_eOya84nPM:&tbnh=160&tbnw=99&usg=vt5Gd0N74Zj_hinkAoAqBxqTSlxg=&yet=10ahUKEwi4nYnMk6HVAhUEpZQKH0zsBTtoQ_B0ldzAK..i&docid=2VylBFdhz7k0M&itg=1&sa=X&ved=0ahUKEwi4nYnMk6HVAhUEpZQKH0zsBTtoQ_B0ldzAK) when he observed that ‘physical concepts are the free creations of the human mind and are not, however it may seem, uniquely determined by the external world’.

In practice, every variable begins with an intention to think of something as varying in amount from lower levels to higher levels, from less to more, shorter to longer, lighter to heavier, colder to hotter – along a single dimension. We invent variables to manage and make sense of complexity. And when we measure, we attempt to estimate locations on the variables we invent: how heavy? how long? how hot?

We recognise that some variables are closely related. For example, height and weight are often highly correlated. But for the purposes of measurement, we keep variables separate in our thinking and attempt to measure them one variable at a time.

In education, the concept of a variable is a fundamental and ubiquitous idea. It is
reflected in our use of words such as ‘better’, ‘deeper’, ‘more’ and ‘higher’; in our description of one student as ‘more able’ or ‘more advanced’ than another; and in references to growth, progress or improvement. All these terms imply the existence of varying amounts of something.

We invoke the idea of a variable when we identify skills, values or attributes that we wish to see students develop, such as ‘a sense of self-worth’, ‘a sense of optimism’, ‘honesty’, ‘resilience’, ‘empathy’ and ‘respect for others’ – all listed as objectives in the Melbourne Declaration on Educational Goals for Young Australians (http://www.curriculum.edu.au/verve/_resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf). In common with other educational variables, these attributes are assumed to exist in varying degrees and to be capable of being developed to a higher level.

We recognise that educational variables, too, are often related. For assessment purposes, however, we keep variables separate in our thinking. For example, we monitor a student’s progress in reading separately from their progress in writing, numeracy or oral language development.

So educational variables are not fundamentally different from other variables. They are creations of the mind, invented to allow us to focus on one aspect of complexity at a time. And each variable reflects an intention to think of something as varying in amount or degree from lower levels to higher levels.

**Instruments**

The role of instruments in measurement is to connect the idea of a variable to real-world experiences. Instruments produce tangible observations that can be used to draw inferences about amounts of an underlying but abstract variable.

For example, a thermometer connects the idea of temperature to the real-world experience of materials expanding when heated. It uses observations of materials expanding as a basis for inferring temperatures, which cannot be observed directly. Similarly, bathroom scales connect the idea of body mass to the fact that springs are compressed by gravitational force. They use observations of springs being compressed to infer weight, which cannot be observed directly.

The construction of instruments is a non-trivial undertaking. It involves identifying the kinds of observations that might be useful for drawing inferences about the variable of interest and then developing practical conditions for making observations. Daniel Fahrenheit spent many years investigating ways of using
expanding materials to infer temperatures.

Although instruments are essential for measurement, in another sense they are unimportant. We do not want measures of our weight to depend on which particular bathroom scales we happen to have used. Measures should have general meanings independent of the instruments used to obtain them.

In education, it is well understood that variables such as reading ability and resilience cannot be observed directly. They must be inferred from observations of how students respond, behave or perform in particular contexts. For example, a student’s level of resilience can only be inferred from observations of their responses to events or situations that demand a level of resilience.

We also understand that the observations we make are not the variables themselves. When we give a child a reading test, we know that they may never again encounter those particular reading passages or those particular questions. That is not the point. The reading passages we use and the questions we ask are merely opportunities to gather relevant observations about what is really of interest – the student’s underlying reading ability, which cannot be observed directly and must be inferred. And because our interest is not in particular passages or questions, we want our inferences about reading ability to transcend the specifics of the passages and questions we happen to have used.

So measuring instruments in education are not fundamentally different from other measuring instruments. Their role is to provide observations that can be used to draw inferences about some variable of interest, and we want our inferences to have general meanings independent of the particulars of the instruments employed.

**Conditions for measurement**

However, not everything we can imagine can be measured.

Consider, for example, the variable ‘creativity’. We might imagine that students differ in their creativity and set about designing ways to measure levels of creativity. But if our assembled observations suggest that being creative in one area tends not to translate to other areas, then the idea of a general ‘creativity' variable is thrown into question, as is our attempt to measure it. An alternative, and perhaps more productive, idea might be that students differ in their creativity within particular areas of activity.

For this reason, every attempt to measure must include a check on the extent to which the collected observations are consistent with the idea of an underlying
variable. This is one role of a measurement model (https://en.wikipedia.org/wiki/Rasch_model) – to supervise attempts to measure to ensure that resulting measures are meaningful. A measurement model provides a mathematical (probabilistic) connection between observations and amounts of an underlying but unobservable variable.

Most attempts to measure are based on observations made in controlled situations. When we measure a person's height, we do not attempt to measure the person as they go about their daily activities, but instead place them in an artificial situation: shoes off, back to the wall, chin up, no slouching. The same is true of measurement in education. But the nature of some variables can make it difficult to make valid observations under standardised conditions. For example, can variables such as honesty, resilience and respect for others be assessed in controlled situations? Will students give the responses they know they should give? And if we gather observations opportunistically in naturally occurring contexts, will these be an adequate basis for measurement?

These challenges are not necessarily insurmountable. They require clever ways of gathering observations coupled with the use of a measurement model to check on the consistency of the observations with the intended variable.

Another significant challenge in measuring educational variables is that the results of attempts to measure are often not comparable across different instruments. A score of 25 on one teacher's mathematics test is unlikely to represent the same level of mathematics proficiency as a score of 25 on another teacher's test. This is similar to every set of bathroom scales producing measures of weight that cannot be compared with weights from any other set of scales. Added to this, a score of 25 questions answered correctly does not have an obvious substantive meaning like 25 centimetres or 25 degrees Celsius.

Again, a measurement model provides the solution. It provides a basis for constructing a measurement scale (with a unit of measurement) against which different instruments can be calibrated. In this way, the scale can be given substantive meaning and measures made with different instruments can be compared directly – essential conditions for the 'measurement' of educational variables.

Whether an educational variable can be measured depends on how creative we are in finding ways to make observations capable of providing valid and consistent information about the variable we are interested in measuring. It also depends on the use of an appropriate measurement model to supervise our efforts to measure.
This article is based on materials developed for ACER’s Graduate Certificate in the Assessment of Student Learning. (https://www.acer.org/professional-learning/postgraduate)