



Reliability

Research Brief

Reliability refers to the consistency of the measurement process and is one of the key criteria for a sound measure of a schooling outcome or school process.

Measures tend to be more reliable when they include a larger number of questions and when there are prescribed methods for collecting data, coding responses, and assigning scores.

Reliability is usually assessed with a coefficient that ranges from 0 to 1.0, with 1.0 being perfectly reliable. Most of the measures used in our surveys have reliability coefficients ranging from 0.80 to 0.95, which are well-suited for guiding school policy and practice.

For measures of non-cognitive student outcomes, such as self-esteem, researchers typically assess the internal consistency of the measure, which gives an indication of how well one can distinguish among members of the sample. Typically, you need at least 5 or 6 well-formulated questions to achieve a measure that is sufficiently accurate to differentiate between individuals. For example, the OurSCHOOL measure of self-esteem is based on six questions and has a reliability coefficient of 0.91. This is quite high and suggests that we can reliably distinguish between students with low versus high levels of self-esteem. It also suggests that if we were to assess a group of students on one occasion, using OurSCHOOL questions, we would be able to reassess them at a later date with the same questions and the results from the two assessments would be highly correlated.

The reliability of an individual student's score is different from the reliability of an aggregate score reported for a school. For measures of students' non-cognitive outcomes, we require measures that are reliable at the individual level, but we are also concerned with how well we can distinguish between schools when it comes to average levels of an outcome or the percentage of students who score above or below a defined threshold.

In most applications of OurSCHOOL, schools use the resulting data to determine how well they fare on certain schooling outcomes, such as students' sense of belonging at school, or aspects of school climate, such as teacher-student relations. It is not used to compare individual students or to identify students who suffer anxiety or depression. Nearly all measures have a school-level reliability above 0.80, which is sufficiently high for school-level policy decisions.

Following the self-esteem example, the school-level reliability coefficient of a school's 'average level of self-esteem' is 0.81. This shows us that schools vary in the average levels of student self-esteem, which suggests we can reliably distinguish between schools with low and high levels of self-esteem. The reliability of a school's score depends on the number of questions and the administrative process, as well as the number of students completing the survey within the school and the extent to which schools naturally vary in a particular outcome. For instance, the OurSCHOOL measure of student anxiety is reasonably reliable at the individual level – 0.84 – but its reliability at the school level is rather low – 0.69. This means that although we can reliably discern which students suffer moderate to severe anxiety, we are less able to determine which schools have a low versus high prevalence of student anxiety.

For measures of certain schooling processes, such as 'teacher-student relations', we are only concerned with the school-level measure of reliability. Indeed, we prefer questions in which students from within a school are consistent in their judgments. For example, the school-level reliability of 'effective learning time' is 0.94, indicating that we can reliably distinguish between schools for this measure.