Validity

Research Brief

Validity refers to whether an instrument is measuring what it is intended to measure. It involves gathering evidence of the fit between the content of the instrument items and the theoretical construct, as well as evidence that the scores on the construct are related to some external criteria.

For example, validating an instrument that purports to measure 'teacher-student relations' would require evidence that the instrument's questions, and the strategy for coding and scoring the responses, led to scores that were consistent with the theoretical definition of teacher-student relations. It would also require collecting evidence about whether the decisions made, based on the instrument, were related to some particular intervention. Recent thinking about validity also requires us to consider the kinds of decisions made from the assessment; a test is valid only if inferences made from the results are appropriate, meaningful, and useful.

The measurement process includes the assignment of numbers to categories of "real-world" observations. Generally, the instrument is a technique for relating what we observe in the real world to some latent or unobserved construct that exists only as part of a theory (Messick, 1989; Wilson, 2005). The latent construct can be a psychological trait such as self-esteem or a measure of classroom learning climate. The measurement process is usually done because we wish to make certain decisions based on people's scores for the underlying construct.

The first step in establishing the validity of an OurSCHOOL measure is to determine a theoretical definition of the construct. This is typically done through a review of the literature, and by seeking expert opinion. The next step is to identify the questions that reflect the defined construct. This process requires thorough discussions with experts, as well as teachers and students in focus groups, and a review of other instruments.

Data are collected for several potential questions with small pilot studies (typically about 50–100 students), and analyzed using two psychometric techniques: factor analysis and item response theory (IRT). The analysis provides an indication of how well the questions hold together to provide a measure of a single unified construct, and whether each question contributes to the reliable measurement of the construct.

The validity of our measures calls for a consideration of how results are interpreted and used in the school context. Measures of schooling outcomes or school processes typically derive their meaning in one of three ways:

1) by comparing a school's results to another standard, such as the state or provincial average;
2) by comparing a school's results to results from other schools; and
3) by tracking a school's performance over time.

The interactive reports for OurSCHOOL Student Surveys use all three approaches to afford meaning to measures of engagement, wellness, and classroom and school climate. The results for a measure can be compared to the national median or average score for middle or secondary schools. The national median, or average score, is usually referred to as 'the norm'. A school can use information about the national norm to establish its own standards.

In some states and provinces, average test scores are reported publicly in 'league tables,' which rank schools on their performance, without any consideration of the ability or family background of students attending the school. Also, for many schooling measures, the average scores of most schools falls in a narrow range in the middle of the distribution. Even a small change in a school's performance can dramatically affect its rank order. Thus, unadjusted comparisons of schools tend to be misleading and unfair to teachers and school administrators.

The Learning Bar views the ranking of schools unfavourably as it is a process that fosters competition rather than discussions of how to best improve student outcomes.

Instead of ranking schools, our scores for each school are compared to a 'replica school'. As soon as a student completes the survey, the score of a replica student is also estimated from the national database. This is accomplished by averaging the scores on a measure for all other students that have the same sex, grade level, parental education, family structure, and level of educational possessions at home. For example, if 600 students in a school complete questions for a measure, then the school's average score for that measure is compared to the average score for 600 replica students. This provides a more reasonable comparison because it takes into account the demographic composition of the school.

Tracking changes in school performance provides a way to assess the effects of school reform efforts that are aimed at improving student outcomes. OurSCHOOL reports can show changes from the beginning to the end of a school year, or from one year to the next.

We strive to increase the validity of our measures by providing workshops and webinars to clients in order to provide the skills necessary to interpret and use the resulting data in appropriate ways. For example, a school with relatively low scores in classroom learning climate might want to focus its professional development efforts on strategies for improving classroom discipline.

We are often asked whether using OurSCHOOL leads to improvements in state or provincial test scores. As with any data monitoring program, the success of OurSCHOOL depends on school staff engaging with the data in

ways that put forward changes in classroom practice and school climate. Results of the Canadian Education Association (CEA) study, What Did You Do in School Today, indicate that all of the schools using OurSCHOOL improved their levels of student engagement. However, the extent of improvement varied among schools, with most of the gains realized after the second year of using OurSCHOOL. The success of the initiative is partly attributable to the efforts of the CEA, Galileo, and The Learning Bar conducting workshops to help schools use the data to effect changes in teachers' classroom practices.