

Understanding Validity of Risk Assessment Instruments

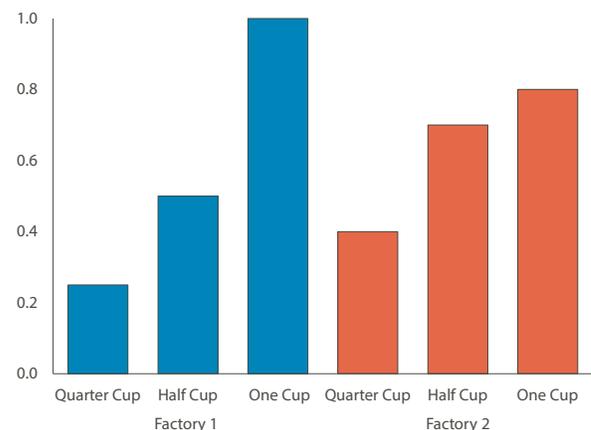
While the abstract concept of validity makes sense, actual testing for validity can be challenging. Because validity exists on a continuum, with degrees of less and more valid, we think of some tools as being more valid than others. This means that a test to determine which tools are most or least valid can be useful.

Validity can be assessed quite simply by looking at a graph depicting how often the tool gets different measures right. In the measuring cup example, we could chart how much flour our one-cup scoop actually gives us. If the measuring cup is valid, then we can expect that most of the time, the one-cup scoop will actually give us one cup of flour. But if the factory that made those measuring cups needs maintenance or the machines need recalibration, then the cups might be off. If your measuring cups aren't valid, the cake you're trying to bake is likely not to turn out how you hope.

What Is Validity?

Validity in general refers to how well an assessment, instrument, or test actually measures the thing it is supposed to be measuring. For example, a measuring cup is generally a good way to measure flour for a cake recipe; you can trust that the amount of flour measured by a measuring cup comes pretty close to a true cup.

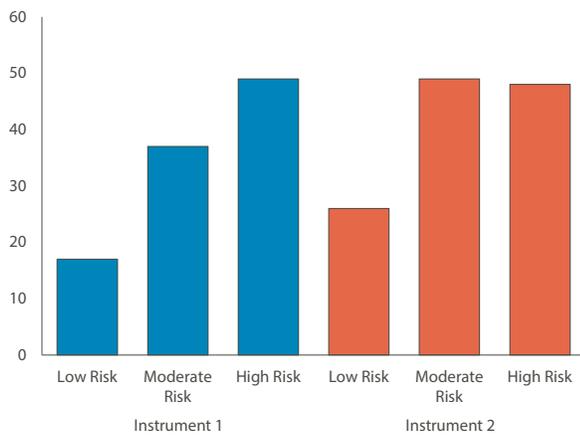
In the figure below, we can see how the measuring cups from Factory 1 (in blue) compare to the measuring cups from Factory 2 (in orange).



The one-cup measure from Factory 1 tends to hold one actual cup of flour, while the one-cup measure from Factory 2 tends to hold less than a cup. The half-cup and quarter-cup measures from Factory 2 also are off. Further, the half-cup does not appear to be very different from the one-cup, holding similar amounts of flour. This chart, without any statistical measure of validity, makes it clear that the measuring cups from Factory 1 are more valid than the measuring cups from Factory 2.

Similarly, if we examine two risk assessment instruments, we can look at recurrence rates by risk level. In the table below, we can compare the

accuracy of classifications made by Instrument 1 (in blue) to the accuracy of classifications made by Instrument 2 (in orange).



Just as it was with the measuring cup factories, it is with the risk assessments. Instrument 1, in this example, shows clear steps up from low risk to moderate risk to high risk. Each level represents a marked increase in recurrence¹ rates. Instrument 2, on the other hand, shows a step up from low risk to moderate risk, but then actually decreases from moderate risk to high risk. (Even for the more valid risk assessments, some low-risk cases have a recurrence and most high-risk cases do not.) This chart, without any statistical measure of validity, makes it clear that the risk levels from Instrument 1 are more valid than the risk levels from Instrument 2.

Some approaches to statistically measuring validity utilize the receiver operating characteristic (ROC) or the area under the ROC curve (AUC). The ROC describes how often a tool produces a

correct result. The AUC is one way to summarize the ROC as a single measure and represents the percentage of randomly drawn pairs for which the test correctly classifies both cases. Other measures, such as the Dispersion Index for Risk, have sought to improve the ability to differentiate between better- and worse-performing tools. **Overreliance on any single measure to describe validity, though, can be problematic. Simple graphical depictions of recurrence rates by risk level can help to confirm the validity of risk assessment instruments.**

It also is worth noting that the accuracy of risk-level classification is not the same as prediction. The goal of a risk assessment is to guide an agency's choice regarding how to allocate resources and target interventions. For more information on the use of risk assessments, please see "[Risk Assessment for Targeting Resources and Interventions.](#)"

In addition, validity is not the only indicator of a good risk assessment. For more information on the other indicators, please see "[Evaluating a Risk Assessment.](#)"



¹ Recurrence refers to a subsequent child protection outcome, such as a referral, substantiation, or investigation.

Find Out More

For more information, please contact research@nccdglobal.org; visit our website, www.nccdglobal.org.