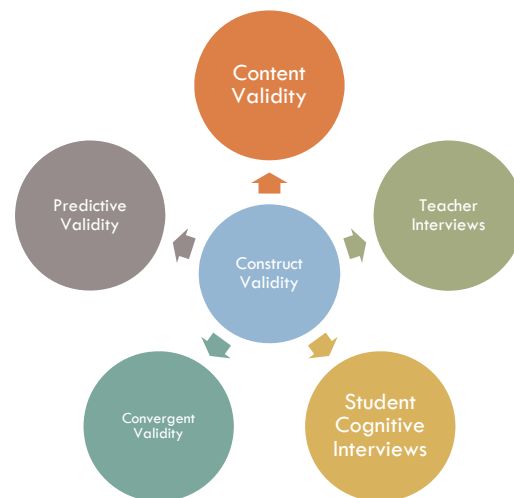


STeP Survey Validation Research Brief

Though we still seek an agreed upon definition of effective teaching, one consistent finding in education research is that there is significant variation among teachers in their ability to increase student achievement. It can be argued that student feedback is an important consideration in any teacher evaluation system as students have the most contact with teachers and are the direct consumers of a teacher's service (Goe, Bell & Little, 2008).

This research brief describes the development and validation of an instrument, known the STeP survey, to measure teacher effectiveness using student feedback. The STeP survey is broken into six different survey categories that represent the various roles that a teacher plays within the classroom. These include Presenter, Manager, Counselor, Coach, Motivator, and Expert.

The validity framework (shown on the right) seeks to establish construct validity through multiple sources of evidence such as content validity, convergent validity, and predictive validity (Messick, 1989). It employs a mixed-method approach to test the survey for its relationship to targeted outcomes as well as internal reliability.



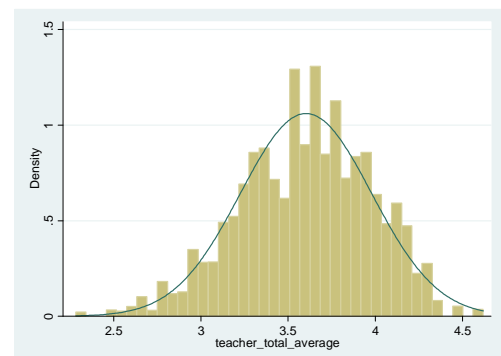
Content validity is established through the development of survey questions. Questions draw upon behaviors that have been consistently identified in the research (such as providing

feedback to students or the use of wait time) as having a positive relationship with academic outcomes. Further, the questions align with validated observation rubrics such as Charlotte Danielson’s Framework for Teaching and CLASS. Both of these procedures allow for the survey to be both research-based and exhaustive of desired teaching behaviors.

Further evidence for construct validity came from cognitive testing with 56 students and 35 teachers to ensure alignment with objectives as well as readability and comprehension. Cognitive interviews were used to determine whether the questions measure what they are intended to measure and the STeP survey is the only survey to have documentation of going through this important step in the validation process. Questions were continually revised and retested to reflect the findings from these interviews.

Following the creation and modification of survey questions, pilot testing represented a way of determining convergent and predictive validity. Pilot testing took place in the spring of 2011, with the majority of work conducted in Georgia as part of the state’s Race to the Top initiative. Overall, 12,408 students completed the survey representing schools from seven districts across the state and 667 teachers. The table and figure below shows the descriptive statistics for a teacher’s overall average.

	Overall
Teacher Total Score	229.92 (25.92)
Teacher Total Average (1-5)	3.63 (0.37)
Number of Students	12,408
Number of Teachers	667



Results demonstrate a positive and significant relationship with all three external measures including value-added student achievement, student academic student engagement, and student academic self-efficacy. The table below shows the results of the correlations with each measure.

	Teacher Total Score	Presenter	Manager	Counselor	Coach	Motivator	Expert
Student Engagement	.675***	.631***	.525***	.612***	.610***	.697***	.548***
Academic Self-Efficacy	.571***	.555***	.419***	.502***	.521***	.550***	.497***
Math Value- Added	.1624*	.1825*	.1491	.1913**	.1529	.1511	.1422
ELA Value- Added	.1917*	.1841*	.1141	.1454	.1668	.1393	.1852*

* p < .10, ** p < .05, *** p < .001

To test the existing scales of the survey, measures of reliability and a confirmatory factor analysis was performed. Generally, it is preferable to have scales with Cronbach Alpha scores of greater than .7, an RMSEA of lower than .1, and a CFI of greater than .9. The table below shows the RMSEA and CFI values for all scales fall within the proper range.

	Presenter	Manager	Counselor	Coach	Motivator	Expert
Cronbach Alpha	0.893	0.704	0.821	0.824	0.850	0.820
RMSEA	.0483	.0599	.0941	.0646	.0995	.0503
CFI	.9758	.9631	.9408	.9250	.9127	.9735
Number of Items	12	6	8	12	9	8

Overall, the STeP survey establishes evidence of construct validity through sources of evidence such as content validity, student and teacher interviews, convergent validity, and predictive validity in addition to high levels of reliability for survey sub-scales.