

# Technical Report: 2020 Teaching and Learning Conditions in Colorado (TLCC) Survey

*Reliability analysis of 2018 to 2020 administrations*

## Overview

Colorado statute requires the Colorado Department of Education (CDE) to administer the Teaching and Learning Conditions Colorado (TLCC) survey to K-12 teachers and principals statewide biennially. For summary reports for the 2018 and 2020 TLCC survey administrations go to [www.TLCCSurvey.org](http://www.TLCCSurvey.org). CDE led the development of the TLCC survey instrument using a process that involved numerous educational leaders, practitioners, researchers, and representatives of education-related organizations in Colorado.

The independent analysis of the reliability and validity of the TLCC survey presented here was conducted by Dr. Kent Seidel, an associate professor of research and evaluation methods and founding director of the Center for Practice Engaged Education Research (C-PEER) at CU Denver.

A similar 2018 technical analysis conducted by Dr. Seidel focused on confirmation of the hypothesized groupings of questions (e.g., *School Leadership*, *Time*, etc.) and whether these items grouped empirically as strong constructs (The technical analysis report is also provided at [www.TLCCSurvey.org](http://www.TLCCSurvey.org)). That report also included missing data diagnosis and comparison of key *overall reflection* survey items with the prior statewide survey, *TELL* (Teaching, Empowering, Leading and Learning) *Colorado* to test the viability of trend comparisons from prior surveys to the new, Colorado-specific TLCC survey.

Since the 2018 technical analysis found the TLCC to be a solid instrument, adequate to continue trend findings from the *TELL Colorado* historical record, the focus of the 2020 analysis was the reliability of the TLCC instrument from administration to administration. This study investigated three main analytic questions regarding the reliability and validity of the TLCC instrument, now that two full administrations' data sets are available:

- 1) Does the 2020 instrument have internal consistency and validity?
  - a. Are there any indications of items biased in responses?
  - b. Does the 2020 data sort (in factor analysis) to reflect the designed item groups (i.e., the constructs)?
  - c. What is the internal consistency (Cronbach's alpha) of the constructs?
- 2) Does the 2020 survey perform in approximately the same fashion as the 2018 survey?
  - a. What is the approximated test-retest reliability?
  - b. How do the construct groupings compare with those in the 2018 administration?
- 3) How do the new survey items added in 2020 and the items with wording modified from the 2018 survey perform? (*See Appendix A for a side-by-side list of 2018 and 2020 items.*)
  - a. Do the 2020 questions with adjusted wording appear to be at least as stable as 2018?
  - b. How do the new questions connect with the construct groupings?

## Data analyzed.

The online 2020 TLCC Survey results include an overall valid number (n) of 37,536 respondents (all classification categories) across 1,958 schools that met the required 50% or higher response rate threshold. The study used the full dataset of teachers and other education professionals, which included responses from schools that were lower than the 50% threshold, to analyze missing data patterns and estimate whether response rate levels affected the validity of the survey.

The 50% or greater threshold for reporting is a way to protect the identity of respondents; the larger data set for the technical analysis contains no identifying information; under a data sharing agreement with the Colorado Department of Education the data are only viewed by Dr. Seidel. Data from schools not meeting the 50% threshold are only used for purposes of technical reliability and validity analysis.

After cleaning, the full technical analysis dataset included 31,869 valid teacher responses and 3,844 other education professional responses. Because leader responses are generally fewer per school (often just 1 or 2 individuals) and a much smaller dataset overall (cleaned = 1,320 valid n), the leader set was not used for this technical analysis of the 2020 survey.

## Internal validity of TLCC instrument, per 2020 data.

- 1) Does the 2020 instrument have internal consistency and validity?
  - a. Are there any indications of items biased in responses?
  - b. Does the 2020 data sort (in factor analysis) to reflect the designed item groups (i.e., the constructs)?
  - c. What is the internal consistency (Cronbach's alpha) of the constructs?

## Item level missing data analysis.

First, the educators' dataset was analyzed to confirm that the 2020 administration did not have relevant missing data biases. There were no correlations between any TLCC construct questions (the items reported by CDE in the public results) and any of the background identifier questions (e.g., role in school, years in position, etc.).

Then, school-level characteristics were connected to the TLCC dataset, and aggregate responses for each school were tested for missing response sets, to see if any particular school types (e.g., size, setting, level, SPF performance rating, charter status - see full list below) were missing in large numbers or non-random patterns. For all construct questions, response option 5 (*I don't know*), was treated as user-missing (intentionally missing) data, and absent responses were identified as system-missing data. The conclusion of the missing data analysis is that missing data in the full educators' dataset are *Missing At Random* for purposes of further analysis, and that the dataset in general is representative of the state-level population at the teacher and school levels.

In addition to item level missing data analysis, a second dataset of school-level educator results was created, with school-level means and standard deviations calculated for all schools meeting 10% or greater response rates (valid n = 1,500 schools). All school-level item results were analyzed in a bivariate correlation table with the school response rates.

While a few statistically significant correlations were identified ( $p \leq .05$ ), the correlation sizes were trivial ( $r^2 < .001$ ). Statistically significant but trivial correlations are often identified in such large datasets. Therefore, the findings suggest that the response rates do not seem to significantly affect school results *overall*, however, it is still a good rule of thumb to use a higher response rate threshold.

While many results are published with low response rates, 60% or greater is considered a good target, with 80% is considered a threshold for many elite peer-reviewed research journals (see US NIH, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2384218/>). Researchers also consider the number of responses, generally targeting 400 as a strong data set as long as the data do not seem to be biased.

For the remaining analyses to address Questions 1a and 1b, the full dataset of responses from schools, removing the very lowest (<10%) response rate sites is used.

### **Construct analysis – 2020 data**

A factor analysis was conducted to test whether the planned constructs, consisting of questions grouped into categories, reflected these categories in the responses of the teachers that completed the TLCC 2020 survey.

Eight categories were tested:

- School Leadership (10 items, 2 new, 4 adjusted wording from 2018)
- Teacher Leadership (4 items, 1 adjusted wording)
- Managing Student Conduct (5 items, 1 new, 1 adjusted wording)
- Instructional Practices & Support (13 items, 3 adjusted wording)
- Professional Development (8 items, 1 new, 4 adjusted wording)
- Time (7 items, 1 new, 5 adjusted wording)
- Facilities & Resources (4 items, 2 adjusted wording)
- Community Support & Involvement (4 items, 2 adjusted wording)
- Overall Reflection (3 items, 1 new)

The groups of questions categorized as *New Teacher Supports* and *District Supports* were not included in the factor analysis or missing data diagnosis, because these questions were only administered to a subset of teachers and to the leaders, respectively.

The factor analysis (Principal Components Analysis, varimax rotation) showed that TLCC 2020 is, empirically, a fairly good match for the constructs as designed, as well as with those constructs as analyzed for the 2018 administration. The analysis shows that, for the most part, the survey groupings as designed reflect the ways that educators throughout Colorado think about these important topics.

However, there are some items that seem to indicate another sub-construct. The *School Leadership* category broke into two factors, as in 2018. In addition, the new *Overall Reflection* item (“I feel satisfied with the recognition I get for doing a good job”) fits best with the recommended new *Teacher Evaluation* grouping.

In addition to the factor analysis loading, removing 3 items from the *School Leadership* grouping increases its alpha from .898 to .922 (7 items), and those 3 items with the addition of the fourth, from *Overall Reflection*, results in a scale alpha of .927.

In consideration of the factor analysis findings, a new *Teacher Evaluation* construct is recommended, which would be comprised of the following:

From the *School Leadership* construct:

- Teachers' effectiveness is accurately assessed through the school's teacher evaluation process.
- The teacher evaluation process provides teachers with actionable feedback for improvement.
- Teachers are provided with informal feedback to improve their instruction.

And from the *Overall Reflection* construct:

- I feel satisfied with the recognition I get for doing a good job.

With one exception, all constructs reached an internal consistency Cronbach's alpha statistic of at least .805 (*Community Support and Involvement*) to a high of .927 (the proposed new *Teacher Evaluation* group). The exception was *Facilities and Resources* (4 items) which only reached alpha = .696. However, the alpha did not increase with deletion of any of the four items, and these items group together in the factor analysis. Because of this, even though the alpha is considered low for research purposes, there is no compelling alternative and the construct should remain as is.

## Comparison of stability of TLCC, 2018 to 2020 administrations.

- 2) Does the 2020 survey perform in approximately the same fashion as the 2018 survey?
  - a. What is the approximated test-retest reliability?
  - b. How do the construct groupings compare with those in the 2018 administration?

One common way to analyze the stability of a survey instrument or assessment is to conduct a test-retest analysis, to determine if the instrument performs in a similar fashion when used a second time. Typically, a test-retest approach would use two administrations of the instrument to the same set of respondents.

While it is not possible to conduct such an administration design, it was possible to approximate it by creating a school-level dataset, and utilizing data (i.e., the school-level means and standard deviations for each TLCC item reported) for schools that had a report in both 2018 and 2020. To further increase the likelihood of similar administrations, the analysis used only schools with 80% or greater response rates in both 2018 and 2020. This resulted in a working dataset of 305 schools, which also represent the variety of demographic variables (e.g., school level, charter/not, and school location).

Although several sources of variability between 2018 and 2020 administrations are evident (i.e., staff changes, or changes in the educational contexts on which the items focus, or that even at 80% response rates, one cannot know if the same 80% of educators in a school are represented in both years), this approximation dataset was assessed for differences between 2018 and 2020 results.

The results from each year were compared using Cohen's d calculation (the difference in means for the item, by school, divided by the pooled standard deviations for that item) to represent the effect size of any change from 2018 to 2020 results. Across all items, for all schools, some schools had very large changes on some items (both to the positive and to the negative), but averaged across the schools by item, only a few items showed even a small effect size (Cohen's d is interpreted as 0.2=small, 0.5=moderate, and 0.8=large). The average effect size across all 58 items, including the non-trivial items noted just below, is only 0.07. These items, even with wording slightly adjusted for 2020 from 2018, are very stable.

The few items that had an effect size (ES) difference across the two administrations large enough to note are listed below. **Readers are strongly cautioned NOT to assume that a general change in direction from 2018 to 2020 administrations indicates a statewide trend.** There are many potential sources of these changes, as noted above, and with one exception, all of these findings only reach what would be considered a small effect size. For example, an ES of 0.2 in an experiment would mean only about 8% difference between those in the experimental group compared to those in the control group (For an overview of ES refer to <https://www.simplypsychology.org/effect-size.html>).

- *My effectiveness is accurately assessed through the school's teacher evaluation process.* (ES 0.26, generally increased in 2020 from 2018).
- *The teacher evaluation process provides me with actionable feedback for improvement.* (ES 0.30, generally increased in 2020 from 2018).
- *The school provides strategies that families can use at home to support their children's learning.* (0.62 ES, generally increased in 2020 from 2018).
- *The community is supportive of the school.* (0.22 ES, generally lower in 2020 from 2018).
- *All families have access to information about what is happening in the school.* (0.24 ES, generally lower in 2020 from 2018).
- *New initiatives (e.g., curriculum, assessments, instructional approach) are given enough time to determine their effectiveness.* (ES 0.29, generally increased in 2020 from 2018).

### Analysis of adjusted item wording and new 2020 items.

- 3) How do the new survey items added in 2020, and the items with wording modified from the 2018 survey perform? (See Appendix A for a side-by-side list of 2018 and 2020 items.)
  - a. Do the 2020 questions with adjusted wording appear to be at least as stable as 2018?
  - b. How do the new questions connect with the construct groupings?

To examine whether the adjusted wording on several 2020 survey items (23 or 43% of items) resulted in substantially different response patterns, the school-level means and standard deviations were used. The dataset was expanded to include schools with greater than 10% response rate that had both 2018 and 2020 survey data. For this item-level analysis, the overall difference of means and the mean difference of standard deviations for each survey question that had been modified were compared. The theory is that if survey items are from the 2018 to 2020 administrations (in spite of slight edits), then the general patterns of means and standard deviations for item responses should not change significantly.







TIME	
I have adequate time to prepare for instruction.	Teachers have adequate time to prepare for instruction.
My time is protected from duties that take time away from teaching.	Teachers' time is protected from duties that take time away from teaching.
I have adequate time to analyze and respond to student assessment data.	Teachers have adequate time to analyze and respond to student assessment data.
I have adequate time to support my students' social and emotional learning.	Teachers have adequate time to support their students' social and emotional learning.
I have adequate time to communicate with my students' families.	Teachers have adequate time to communicate with their students' families.
New initiatives (e.g., curriculum, assessments, instructional approach) are given enough time to determine their effectiveness.	New initiatives (e.g., curriculum, assessments, instructional approach) are given enough time to determine their effectiveness.
	Teachers have adequate time to engage in professional learning (e.g., attend trainings, refine new techniques, collaborate with grade level teams).
FACILITIES AND RESOURCES	
My class size(s) is reasonable.	Class size(s) are reasonable.
Instructional resources are adequate to support student learning.	Instructional resources are adequate to support student learning.
I have adequate physical space to work productively.	Teachers have adequate physical space to work productively.
Our school is a safe place to work.	Our school is a safe place to work.
COMMUNITY SUPPORT AND INVOLVEMENT	
The community is supportive of the school.	The broader community is supportive of the school.
The school's efforts to engage families are effective.	The school's efforts to engage families are effective.
The school provides strategies that families can use at home to support their children's learning.	The school provides strategies that families can use at home to support their children's learning.
All families have access to information about what is happening in the school.	Every family has access to information about what is happening in the school.
OVERALL REFLECTION	
I would recommend this school as a good place to work.	I would recommend this school as a good place to work.
I would recommend this school as a good place for students to learn.	I would recommend this school as a good place for students to learn.
	I feel satisfied with the recognition I get for doing a good job.