Technical Brief: Cognia Teacher Observation Tool

V1.0, May 8, 2023

1. Introduction

The Cognia® Teacher Observation Tool is a formative observation tool designed to foster teachers' professional improvement toward the practice of learner-centric teaching. This observation tool provides rich formative data for teachers and administrators to engage in clear and focused discussions about improving teaching practices based on information gathered from direct classroom observations. Accompanying this tool is the Teacher Observation Tool Rating Guide. By rating the items within the five overarching dimensions related to high-quality instruction, administrators and teachers can identify crucial elements to help strengthen and sustain effective teaching practices.

2. Background

The design of Cognia's Teacher Observation Tool involved an iterative process, meaning that several phases of development design were executed before the content was determined ready for field testing. The process began with an extensive environmental scan to understand current teacher observation and evaluation practices and resources used in regions in the United States and globally. These findings, once analyzed, drove the next phase of the development process, the researching phase. In this phase, key research questions were established, literature reviewed, and theoretical constructs formed that led to the eventual development of the five dimensions that frame the Teacher Observation Tool. Considerable reading in research disciplines such as teacher effectiveness, motivational theory, classroom equity, and learner-centric instruction served as the foundation for the beginning of the tool's conceptual design and in its final review. Aligning with Cognia's beliefs regarding learner-centric teaching and student engagement, learning as defined by Palincsar (1998) grounded the initial discussions and readings relative to creating a formative observation tool: "Learning is the process of actively engaging learners in building knowledge, whereas instruction is the process by which teachers intentionally and purposefully engage learners in said knowledge-building." Furthermore, research on effective teaching indicated that effective teachers are characterized by qualities related to the following categories of effective teaching: content knowledge, pedagogical knowledge, delivery of instruction, assessment of/for learning, learning environment, and professional disposition (Bulger, Mohr, & Walls, 2002; Stronge, 2018; Walker, 2008).

Following the researching phase, the development team identified five dimensions of high-quality instruction to progress to the ideating and creating phases of the development process. The items for each dimension, in draft form, were written to align to the research reviewed in the previous phase, describing high-quality learner-centric practices and an equitable classroom

environment. Each item was carefully written and vetted to ensure that the final pilot test items met the criteria for quality instrument design and reflected high-quality teacher actions. All words were checked by the development team and by a group of domestic and international educators to ensure each word was free of bias, emotionality, and jargon, and was observable and applicable in any classroom setting including remote, digital, and in-person. Additionally, each item was constructed so that it was observable in any K–13 classroom regardless of the subject taught. Creating the dimensions and subsequent items were one component of the creating and ideating phases; an additional tool, a rating guide, like a rubric, went through the same development phases as the Teacher Observation Tool.

The last phase of the development process, prototyping, involved testing the initial form of the Teacher Observation Tool to determine if it could be used as designed and intended (a minimum of 20 minutes to observe all items in any K–13 classroom setting). This small-scale pilot testing was conducted by the development team and its results were used to determine if the Teacher Observation Tool needed further revisions and research (the iterative development process). Together, these phases and processes yielded the final Teacher Observation Tool.

3. Instrument Overview

The Cognia Teacher Observation Tool encompasses effective teaching practices within the following five dimensions: Culture and Climate, Learning, Essentials, Agency, and Relationship.

- The Culture and Climate Dimension describes a classroom environment anchored by learners and teachers mutually respecting, genuinely supporting, and encouraging each other. In this environment, learners have a deep sense of belongingness that is fostered by the teacher modeling high expectations, equity, and inclusivity. Learners feel loved and cared for, safe, and secure to take the social, emotional, and academic risks needed to self-actualize (Barr, 2016; Harbour et al., 2015; Mayne, 2019; Titsworth et al., 2013).

- The Learning Dimension describes the processes established by the teacher to ensure learners have opportunities to demonstrate an understanding of the content and apply that knowledge and acquisition of skills. Learning is not something done to learners but is something learners do with the content. It is the end product of how students understand and react to their experiences (Bell et al., 2019; Coe et al., 2014; Hammond-Bennett et al., 2014; Peterson-Deluca, 2016).

- The Essentials Dimension describes the core competencies demonstrated by an effective teacher, including presence and disposition, pedagogical knowledge, content knowledge, and management of the learning environment (Bell et al., 2019; Coe et al., 2014; Kimmel et al., 2019; Pierce, 2019).

- The Agency Dimension describes the actions effective teachers use to foster learners' abilities to self-direct and self-regulate their own learning resulting in a growth mindset that includes a strong sense of intrinsic motivation and self-efficacy. This includes taking the initiative to craft, carry out, and, if needed, revise an actionable plan for improvement while engaged in continuous self- reflection (Bell et al., 2019; Coe et al., 2014; Habour et al., 2015; Muijs et al., 2014).

- The Relationship Dimension describes the actions effective teachers use to foster healthy, positive connections between and among teachers and learners. These relationships are foundational to learners developing their social-emotional well-being and growing their capacity to navigate a global and diverse world (Barr, 2016; Mayne, 2019; Muijs et al., 2014; Peterson-Deluca, 2016).

4. Administration

Cognia offers year-round access to the Teacher Observation Tool, and members can choose to administer the observation tool at any point they consider appropriate. To date, over 52,000 observations across the world have been conducted. In the 2022–23 administration, the distribution of administration dates differed between U.S. schools and international schools. Observations conducted in the U.S. schools continued to increase from the launch date. The number of observations per week stabilized in November, except holidays, with slight increases week over week, as shown in Figure 1. International schools had fewer observations administered than U.S. schools, but their pattern was similar to that of U.S. schools, as shown in Figure 2.

Figure 1: Distribution of Observation Responses Throughout the Year for U.S. Schools
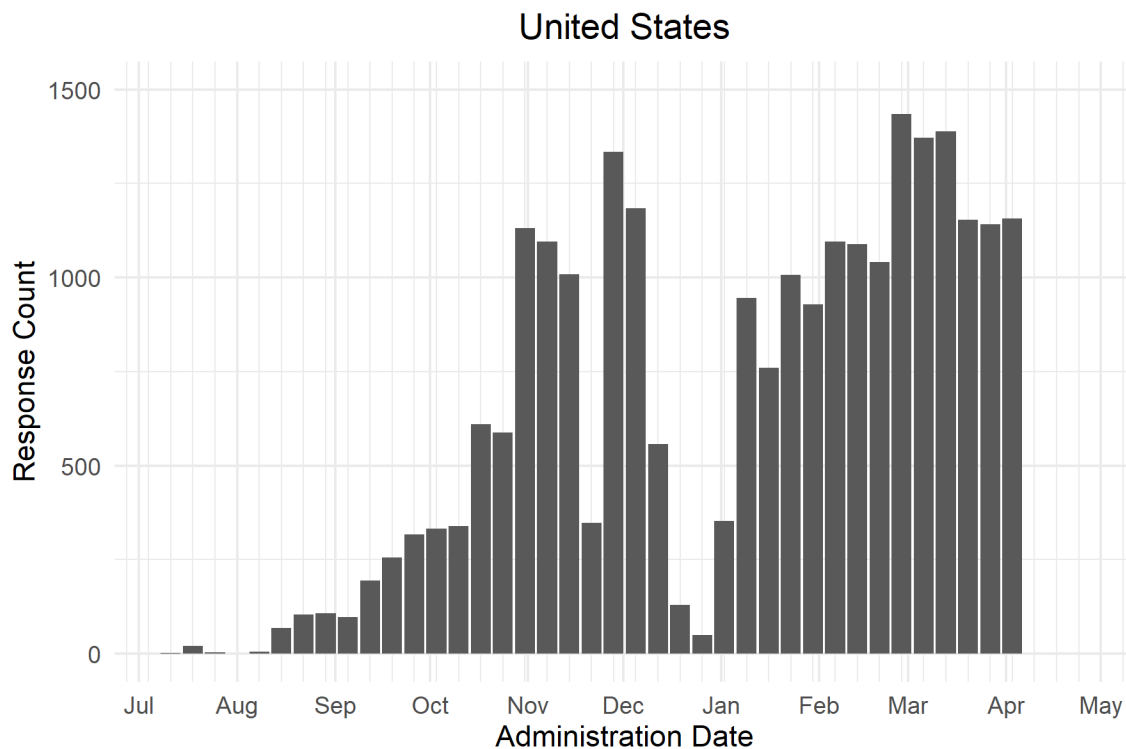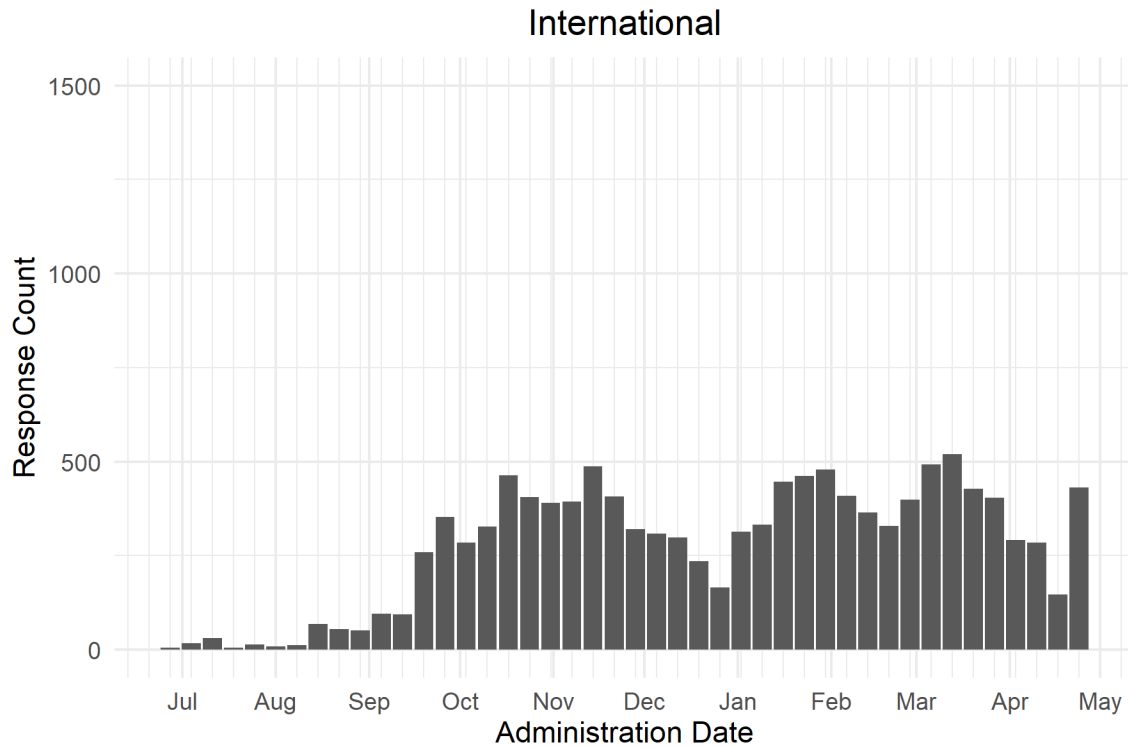
Figure 2: Distribution of Observation Responses Throughout the Year for International Schools

5. Participation

A total of 42,603 observations from 2,127 institutions were conducted by 4,082 observers from July 1, 2022 through April 30, 2023. The results described in this report are based on completed observations in grade K through grade 13.

Participation by Geographical Regions

These observations were conducted by 2,127 educational institutions globally. Institutions from the United States comprised 82.8% of the total number of institutions and administered 71.6% of the total observations. The remaining 17.2% of institutions were international and administered 28.4% of the total observations. Among the U.S. institutions, most observations were concentrated in the southeast region (Table 1).

TABLE 1. Participation by Geographical Regions

	Observations		Observers		Institutions	
Region	n	%	n	%	n	%
International	12090	28.4	1282	31.4	366	17.2

Technical Brief

Mid-Atlantic	2128	5.0	339	8.3	218	10.2
Midwest	1006	2.4	124	3.0	86	4.0
Mountain	967	2.3	139	3.4	108	5.1
Northeast	281	0.7	30	0.7	31	1.5
Pacific	730	1.7	80	2.0	43	2.0
Southeast	25401	59.6	2088	51.2	1275	59.9
Overall	42603	100	4082	100	2127	100

Participation by Grades

All grade levels from grade K through grade 13 were represented in the 2022–23 administration. An observation can cover more than one grade when the classroom has students from mixed grades. The number of observations was nearly evenly distributed across grades with a slightly higher number in grades 10, K, 1, and 11. By grade span, 2,506 schools have early elementary grades (K–2), 2,531 schools have elementary grades (3–5), 2,023 schools have middle school grades (6–8), and 2,654 schools have high school grades (9 or greater) (Table 2).

TABLE 2. Participation by Grades

	Observations		Observers		Institutions	
Grade	n	%	n	%	n	%
K	4111	8.8	1077	7.3	807	8.3
1	3870	8.3	1240	8.4	866	8.9
2	3722	8.0	1229	8.3	833	8.6
3	3760	8.1	1265	8.5	869	8.9
4	3405	7.3	1257	8.5	845	8.7
5	3266	7.0	1191	8.0	817	8.4
6	3074	6.6	1078	7.3	684	7.0
7	3327	7.2	1102	7.4	673	6.9
8	3352	7.2	1081	7.3	666	6.9
9	3676	7.9	1148	7.7	679	7.0
10	4177	9.0	1152	7.8	679	7.0
11	3791	8.1	1068	7.2	657	6.8
12	2806	6.0	875	5.9	565	5.8
13	185	0.4	80	0.5	74	0.8
Overall	46522	100	14843	100	9714	100

6. Descriptive Analysis

The Teacher Observation Tool consists of items categorized by five dimensions of observations: Culture and Climate Dimension (items A1–A4), Learning Dimension (items B1–B6), Essentials

Dimension (items C1–C5), Agency Dimension (items D1–D5), and Relationship Dimension (items E1–E3). Point values from 1 to 4 correspond to four evidence levels: Not Evident, Somewhat Evident, Evident, and Very Evident.

6.1 Item Analysis

Tables 5 provides the item statistics, including means (average response to an item), standard deviations ("sd"; dispersion of scores; how "spread out" the scores are for that item), item correlations ("corr w total"; relationship between item score and total scores), and answer option percentage (proportion of each answer option). Figure 3 displays the proportion of each answer option by item. The item text for each dimension can be found in Appendix A.

Notable Findings

- Item analysis showed mean scores of items ranging from 2.56 (Item 2 in Agency Dimension) to 3.54 (Item 1 in Relationship Dimension). The second highest item mean is 3.52 (Item 1 in Essentials Dimension).
- The median value of item means is 3.25. The mean value of item means is 3.22. Standard deviation values were clustered relatively closely to the means (0.62 to 1.03).
- Item score to total score correlations ranged from 0.58 to 0.77. The lowest correlations were present in the Agency Dimension.
- Item 1 in Relationship Dimension (Item E1), Item 1 in Essentials Dimension (Item C1), and Item 3 in the Relationship Dimension (Item E3) had the highest proportion of "Very Evident," 60.5%, 59.8%, and 55.5% respectively.
- Item 2 in the Agency Dimension observed the highest proportion of "Not Evident" at 20.9%. Item 2 in the Culture and Climate Dimension had the lowest proportion of "Not Evident," at less than 1%.

TABLE 5. Item Descriptive Statistics

Dimension	Item ID	n	mean	sd	corr w total	Answer Option Percentage			
						p1	p2	p3	p4
Culture/Climate	A1	42603	3.42	0.66	0.74	1.0	7.0	41.1	50.9
	A2	42603	3.48	0.62	0.71	0.7	4.7	40.3	54.2
	A3	42603	3.34	0.76	0.68	3.4	7.7	40.9	48.0
	A4	42603	3.30	0.75	0.73	2.0	11.3	41.5	45.2
Learning	B1	42603	3.35	0.72	0.72	1.8	9.0	41.8	47.4
	B2	42603	3.06	0.82	0.71	4.5	17.4	45.8	32.2
	B3	42603	3.12	0.79	0.68	4.3	13.3	48.3	34.1
	B4	42603	3.25	0.73	0.74	2.0	11.2	46.2	40.6
	B5	42603	2.94	0.90	0.70	8.6	17.8	44.2	29.4
	B6	42603	3.16	0.80	0.74	4.1	13.2	45.4	37.3
Essentials	C1	42603	3.52	0.65	0.70	1.3	5.0	33.9	59.8

	C2	42603	3.27	0.74	0.75	2.2	11.3	44.3	42.3
	C3	42603	3.21	0.77	0.71	3.1	11.8	46.0	39.0
	C4	42603	3.18	0.79	0.77	3.2	14.2	44.2	38.3
	C5	42603	3.32	0.74	0.73	2.0	10.5	41.1	46.3
	D1	42603	3.20	0.74	0.76	2.6	11.8	48.3	37.3
	D2	42603	2.56	1.03	0.58	20.9	22.1	37.5	19.6
Agency	D3	42603	3.04	0.86	0.69	6.8	14.6	46.5	32.1
	D4	42603	3.06	0.84	0.74	5.8	14.6	47.1	32.4
	D5	42603	3.04	0.84	0.76	6.3	15.0	47.5	31.2
	E1	42603	3.54	0.63	0.69	1.0	4.5	34.1	60.5
Relationship	E2	42603	3.31	0.81	0.70	4.3	9.2	37.8	48.8
	E3	42603	3.47	0.68	0.70	1.8	5.1	37.6	55.5

Figure 3. Percentage of Each Answer Options by Item
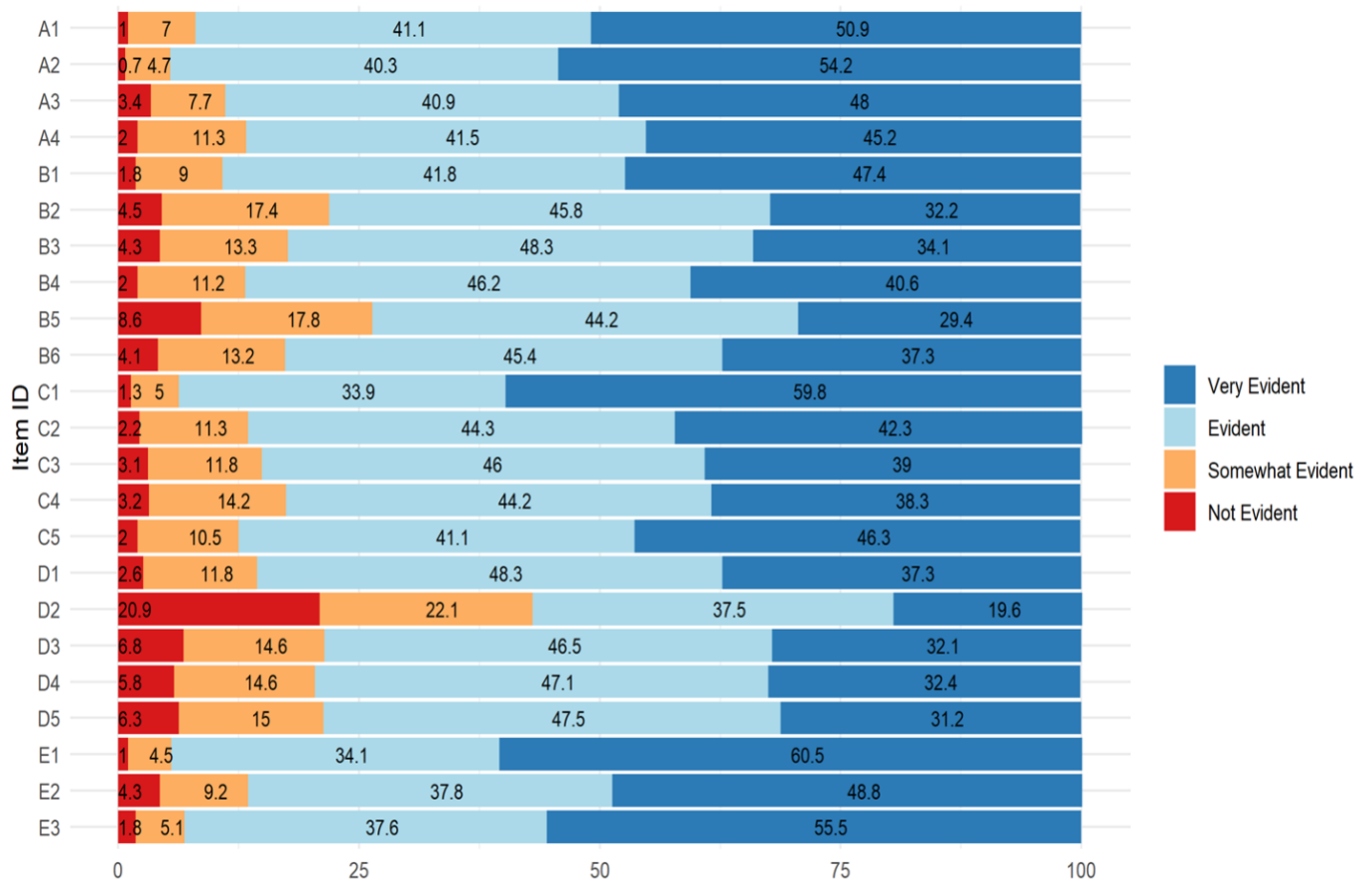
6.2 Dimension Analysis

Dimension analysis provides an overview of dimension performances by aggregating items to their parent dimension level. Within each of the five dimensions, there are higher and lower scored items, however all dimensions demonstrated sufficient mean scores.

Notable Findings

- Mean scores of the dimensions ranged from 2.98 to 3.44. The Relationship Dimension and the Culture and Climate Dimension had the highest mean scores, while the Agency Dimension had the lowest mean score and the highest standard deviation.
- All dimensions were highly correlated with total scores. The correlation ranged from 0.81 to 0.94.
- Items in the Agency Dimension also showed the highest proportion of "Not Evident" among all dimensions.
- Institutions performed the best in the Relationship Dimension and Culture/Climate Dimension. Items in these two dimensions had high mean scores, low standard deviations, and high proportions of "Evident" and "Very Evident."

Table 6 shows descriptive statistics for each dimension.

TABLE 6. Dimension Descriptive Statistics

Dimension	n	mean	sd	min	max	corr w total
A. Culture/Climate	42603	3.40	0.57	1	4	0.87
B. Learning	42603	3.15	0.61	1	4	0.94
C. Essentials	42603	3.30	0.59	1	4	0.92
D. Agency	42603	2.98	0.68	1	4	0.89
E. Relationship	42603	3.44	0.60	1	4	0.81

7. Reliability and Validity

7.1 Internal Consistency

Observation responses were used to investigate the evidence for the internal consistency reliability across the instrument. Reliability was calculated according to Cronbach's Alpha (α) and McDonald's omega (ω)[1]. The Teacher Observation Tool demonstrates α values of 0.955, and ω values of 0.976. These values provide evidence that the tool demonstrates sufficient reliability

---

[1] Though Cronbach's Alpha is widely used and understood as a measure of reliability, McDonald's omega is a distinct reliability coefficient with the advantage of considering the strength of association between items, leading to a stronger measurement of the scale's internal consistency.

to make decisions about effective teaching practices within the dimensions[2]. Table 7 provides reliability for the whole instrument and each dimension.

TABLE 7. Internal Consistency and Reliability of Dimensions

Reliability	A	B	C	D	E	Total
Alpha (α)	0.834	0.854	0.857	0.839	0.807	0.955
Omega (ω)	0.863	0.883	0.871	0.858	0.820	0.976

A. Culture/Climate Dimension; B. Learning Dimension; C. Essentials Dimension; D. Agency Dimension; E. Relationship Dimension

7.2 Validity Evidence

Validity evidence for the measurement of the Teacher Observation Tool is presented in Table 8 as correlations between dimensions for the whole instrument. Specifically, correlations of responses across dimensions are presented in the off-diagonal cells and represent discriminant validity evidence (indicating the degree to which dimensions are separate and distinct from one another), while convergent validity evidence is presented in the cells along the diagonal as reliability (i.e., ω; indicating the consistency of measurement within the dimensions). Within-dimension reliabilities range from 0.820 to 0.883, demonstrating sufficient and strong convergent validity evidence. Across-dimension correlations are moderate to strong and range from 0.652 to 0.835; however, the across-dimension reliabilities (i.e., discriminant validity evidence) are consistently lower than the associated within-dimension reliabilities. These results indicate that the Teacher Observation Tool consistently measures effective teaching practices within each of the five dimensions, which are related to each other, but still demonstrates sufficient independence as to be considered separate traits or constructs.

TABLE 8. Convergent and Discriminant Validity of Dimensions

Dimensions	Culture/Climate	Learning	Essentials	Agency	Relationship
Culture/Climate	0.863				
Learning	0.760	**0.883**			
Essentials	0.782	0.835	**0.871**		
Agency	0.677	0.814	0.740	**0.858**	

---

[2] Salvia, J., & Ysseldyke, J. E. (2004). Assessment in special and inclusive education (9th ed.). Boston: Houghton Mifflin.

Technical Brief

Relationship	0.744	0.686	0.708	0.650	**0.820**

7.3 Construct Validity

The fit of the theoretical and conceptual structure of the Teacher Observation Tool to data resulting from classroom observations can be evaluated through statistical techniques known as latent factor models. Specifically, two latent factor models can be used to describe the relationship among items, dimensions (A. Culture/Climate Dimension; B. Learning Dimension; C. Essentials Dimension; D. Agency Dimension; E. Relationship Dimension), and overall instructional effectiveness as captured by the Teacher Observation Tool. The Unidimensional Model conceives of observed ratings as being correlated to overall performance; the First-Order Model suggests that observed ratings are each correlated with a specific dimension and these dimensions are correlated with each other.

Confirmatory Factor Analysis (CFA) is the statistical technique used to evaluate whether the Unidimensional Model or the First-Order Model fits the data best and to confirm that the target model fits the data well. Model comparison is conducted according to a Likelihood-Ratio Test, with significant differences indicating that the more complex (i.e., First-Order) model fits best; model evaluation is conducted by comparing standard model fit statistics (RMSEA and χ^2/df) against critical values in an empirical sampling distribution that was generated according to a bootstrap resampling procedure (replications = 100; $\alpha = 0.05$).

The model evaluation found that RMSEA and χ^2/df values did not exceed the empirically determined significance values; therefore, the models all fit the data sufficiently well to enable further comparison and interpretation[3]. Model comparison according to the results of Likelihood-Ratio Tests between the Unidimensional Model and First-Order Model found that First-Order Model fits significantly better than the Unidimensional Model. Together, these results provide evidence supporting measurement and reporting of the Cognia Teacher Observation Tool results according to the five teaching dimensions. Table 9 summarizes model evaluation and comparison results.

TABLE 9. Confirmatory Factor Analysis Model Summary

Model	N	df	RMSEA	χ2/df	Notes
Unidimensional	42,603	230	0.060	154.294	

---

[3] Kline, R. B. (2013). Beyond significance testing: Statistics reform in the behavioral sciences, 2nd Edition. Washington, D.C.: American Psychological Association.

First-Order	42,603	220	0.041	72.628	*

* Preferred model based on LRT results

For Additional Information

For any additional questions about this study contact the authors:

Xiaonan Zhang, Measurement Scientist: xiaonan.zhang@cognia.org

Monica Blomker, Director of Effectiveness: monica.blomker@cognia.org

Appendix A

Table A1: Cognia Teacher Observation Tool Instrument Item Text

Dimension	Item	Item Text
Culture/Climate	A1	Fosters an environment that embraces all learners
	A2	Treats each learner equitably
	A3	Encourages learners to share their opinions without fear of negative comments from their peers
	A4	Creates enthusiasm for the learning at hand
Learning	B1	Communicates clear explanations about the activities or tasks
	B2	Implements lessons and/or activities that stimulate learners to use higher order thinking skills
	B3	Delivers lessons that are relatable to the learners or aligned to their interests
	B4	Monitors learners' understanding of the content and/or the acquisition of skills
	B5	Adapts instruction and/or activities that meet individual learner's needs
	B6	Provides learners with purposeful feedback about their progress and/or needs
Essentials	C1	Delivers and/or facilitates the lesson with knowledge and confidence
	C2	Communicates and upholds high expectations for learners' behaviors to maximize their learning and well-being
	C3	Facilitates use of resources that support learners' needs
	C4	Implements instructional strategies that actively engage learners
	C5	Manages the learning time in an efficient and optimal manner
Agency	D1	Empowers learners to be responsible for the learning at hand

	D2	Gives learners choices about the learning activities or tasks
	D3	Provides assistance for learners to navigate and monitor their learning progress
	D4	Encourages learners to persevere with or seek challenging activities or tasks
	D5	Builds learners' growth mindset and self-efficacy
Relationship	E1	Promotes respectful and caring interactions toward and between learners
	E2	Cultivates learner cooperation, collaboration, and inclusivity
	E3	Preserves learners' dignity while attending to their individual needs

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Barr, J. J. (2016). Developing a positive classroom climate. The IDEA Center. *61*, 1–9. https://www.researchgate.net/publication/312021719

Bell, C. A., Dobbelaer, M. J., Klette, K. & Visscher, A. (2019). Qualities of classroom observation systems. School Effectiveness and School Improvement, *30*(1), 3–29. https://doi.org/10.1080/09243453.2018.1539014

Bonnet, D.G. (2002). Sample size requirements for testing and estimating coefficient alpha. Journal of Educational and Behavioral Statistics, *27*(4), 335–340.

Bulger, S., Mohr, D., & Walls, R. (2002). Stack the deck in favor of your students by using the four aces of effective teaching. https://www.researchgate.net/profile/Derek_Mohr/publication/282135751_Stack_the_ Deck_in_Favor_of_Your_Students_by_Using_the_Four_Aces_of_Effective_Teaching

Cantrell, S., & Kane, T. J. (2013). Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study. MET Project Research Paper.

Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). What makes great teaching? Review of the underpinning research. Project Report. Sutton Trust, London. https://dro.dur.ac.uk/13747/1/13747.pdf

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, *20*, 37–46.

Hammond-Bennett, A., Upton, A., & Mason, S. (2014). Accomplished teachers' implementation of quality teaching practices. Physical Educator, *71*(2), 320–343.

Harbour, K. E., Evanovich, L. L., Sweigart, C. A., & Hughes, L. E. (2015). A brief review of effective teaching practices that maximize student engagement. Preventing School Failure, *59*, 5–13. https://doi.org/10.1080/1045988X.2014.919136

Kimmel, S. C., Mardis, M. A., Pribesh, S., Pasquini, L. A., Schulz-Jones, B., Jones, F. R., Wine, L. D., & Colson, L. M. (2019). The preparation and certification of school librarians: Using causal educational research about teacher characteristics to probe facets of effectiveness. School Library Research, *22*, 32.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. Journal of Psychiatric Research, *45*(5), 626–629. https://doi.org/10.1016/j.jpsychires.2010.10.008

Mayne, H. (2019). Pedagogical content knowledge and social justice pedagogical knowledge: Re-envisioning a model for teacher practice. Research in Educational Administration & Leadership, *4*(3), 701–718. https://doi.org/10.30828/real/2019.3.9

McDonald, R.P. (1999). Test theory: A unified treatment. Hillsdale: Erlbaum.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. School Effectiveness and School Improvement, *25*(2), 231–256. https://doi.org/10.1080/09243453.2014.885451

Palinscar, A. S. (1998) Social constructivist perspectives on teaching and learning, Annual Review of Psychology, *49*, 345–375. Retrieved from https://doi.org/10.1146/annurev.psych.49.1.345

Peterson-DeLuca, A. (2016.) Top five qualities of effective teachers, according to students. Pearson PreK-12 Education. https://www.pearsoned.com/top-five-qualities-effective-teachers/

Pierce, S. (2019). The importance of building collective teacher efficacy: Leadership and accountability are crucial to student success. Leadership, *22*.

Revelle, W. & Zinbarg, R.E. (2009) Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. Psychometrika, *74*, 145.

Rotondi, M.A., & Donner A. (2012). A confidence interval approach to sample size estimation for interobserver agreement studies with multiple raters and outcomes. Journal of Clinical Epidemiology, *65*, 778–784.

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2013). Assessment in special and inclusive education. Cengage Learning.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, *86*(2), 420.

Stronge, J. H. (2018). *Qualities of effective teachers*. Alexandria, VA: ASCD.

Titsworth, S, McKenna, T.P., Mazer, J.P., & Quinlan, M.M. (2013) The bright side of emotion in the classroom: Do teachers' behaviors predict students' enjoyment, hope, and pride? Communication Education, *62*(2), 191–209. https://doi.org/10.1080/03634523.2013.763997

Walker, R. J. (2008). Twelve characteristics of an effective teacher: A longitudinal, qualitative, quasi-research study of in-service and pre-service teachers' opinions. Educational Horizons, *87*, 61–68. Retrieved from https://files.eric.ed.gov/fulltext/EJ815372.pdf