

# APPENDIX A

## 2017-2018 CIS PROTOTYPE – PARTICIPATING INSTITUTIONS



**12 Diverse Providers**

**2**

Private

**8**

Public

**2**

Hybrid



**Serving 8,700 candidates**

Undergraduate + Graduate + Residency + Alternative

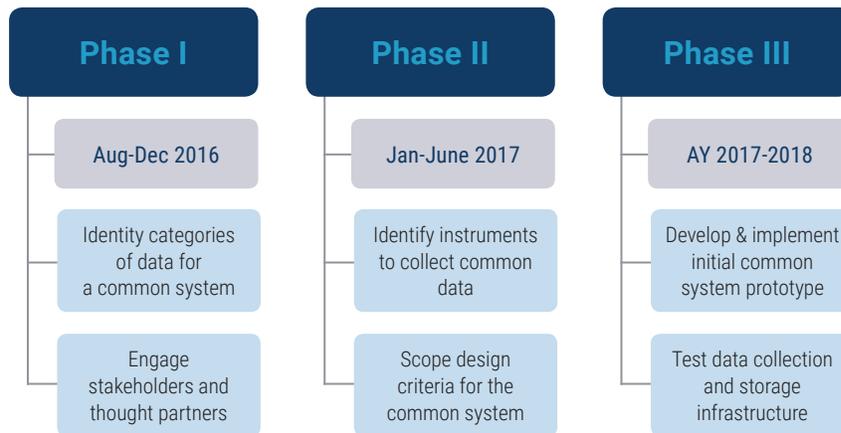
### Institutions Participating in the CIS Network Prototype

- ◆ Arizona State University
- ◆ Boston Teacher Residency
- ◆ Loyola Marymount University
- ◆ Relay Graduate School of Education
- ◆ Temple University
- ◆ Texas Tech University
- ◆ University of Nevada, Reno
- ◆ University of North Carolina at Charlotte
- ◆ University of Southern California
- ◆ University of Texas Rio Grande Valley
- ◆ University of Virginia
- ◆ Urban Teachers

# APPENDIX B

## INSTRUMENT SELECTION PROCESSES

The CIS Network took pains to establish a truly participant-owned process for identifying and testing the measurement instruments at the core of our common data system. That process unfolded in two phases prior to trialing them in a pilot year:

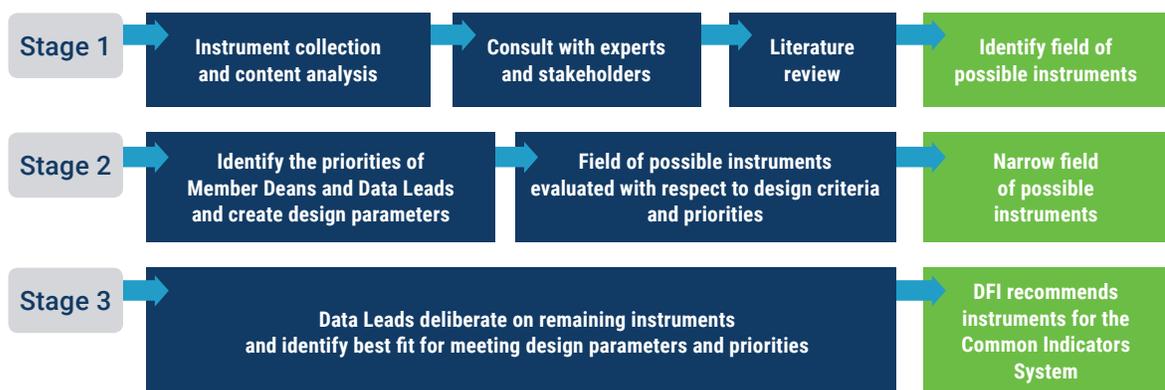


### Phase I

From August to December 2016, member deans refined categories of data for the initial Common Indicators System pilot. They agreed to five categories: Observations of candidate skill, candidate dispositional assessments, a graduate survey, an employer survey, and a model MOU with schools for accessing outcomes data.

### Phase II

Phase II involved intensive analysis to select appropriate instruments to measure the four categories (the toolkit encompassing a model MOU occurred in a separate process).



**Stage 1** began with a systematic review of the instrument landscape where we collected the instruments being used by member institutions in each category as well as those recommended by researchers, practitioners, and other stakeholders. We used the literature review conducted in Phase I to review the research base on each of the identified instruments, and completed a content analysis on each, considering multiple factors:

Observation Rubrics	Dispositional Measures	Graduate Surveys	Employer Surveys
<ul style="list-style-type: none"> <li>◆ # of Rating Categories</li> <li>◆ Rating Descriptors</li> <li>◆ Frequency of Obs.</li> <li>◆ Instructional Domains</li> <li>◆ Instructional Components</li> <li>◆ Reliability Evidence</li> <li>◆ Validity Evidence</li> <li>◆ Evaluator(s)</li> <li>◆ Time of Observations</li> <li>◆ Program Sequence</li> <li>◆ Training Procedures</li> <li>◆ Alignment to Standards</li> <li>◆ Summative Use</li> <li>◆ Formative Use</li> <li>◆ Development Process</li> <li>◆ Adoption Process</li> <li>◆ Years in Use</li> <li>◆ Value of Data to Programs</li> </ul>	<ul style="list-style-type: none"> <li>◆ Form of Assessment</li> <li>◆ # of Rating Categories</li> <li>◆ Rating Descriptors</li> <li>◆ Freq. of Administration</li> <li>◆ Dispositions Assessed</li> <li>◆ Prof. Behaviors Assessed</li> <li>◆ Reliability Evidence</li> <li>◆ Validity Evidence</li> <li>◆ Evaluators</li> <li>◆ Time of Administration</li> <li>◆ Program Sequence</li> <li>◆ Training Procedures</li> <li>◆ Alignment to Standards</li> <li>◆ Summative Use</li> <li>◆ Formative Use</li> <li>◆ Development Process</li> <li>◆ Adoption Process</li> <li>◆ Years in Use</li> <li>◆ Value of Data to Programs</li> </ul>	<ul style="list-style-type: none"> <li>◆ Instructions</li> <li>◆ Stem</li> <li>◆ Topic Areas Covered</li> <li>◆ Reliability Evidence</li> <li>◆ Validity Evidence</li> <li>◆ Scale</li> <li>◆ Descriptors</li> <li>◆ Freq. of Administration</li> <li>◆ Time of Administration</li> <li>◆ Program Sequence</li> <li>◆ Alignment to Standards</li> <li>◆ Formative Use</li> <li>◆ Summative Use</li> <li>◆ Development Process</li> <li>◆ Adoption Process</li> <li>◆ Years in Use</li> <li>◆ Value of Data to Programs</li> </ul>	<ul style="list-style-type: none"> <li>◆ Instructions</li> <li>◆ Stem</li> <li>◆ Topic Areas Covered</li> <li>◆ Reliability Evidence</li> <li>◆ Validity Evidence</li> <li>◆ Scale</li> <li>◆ Descriptors</li> <li>◆ Freq. of Administration</li> <li>◆ Time of Administration</li> <li>◆ Program Sequence</li> <li>◆ Alignment to Standards</li> <li>◆ Formative Use</li> <li>◆ Summative Use</li> <li>◆ Development Process</li> <li>◆ Adoption Process</li> <li>◆ Years in Use</li> <li>◆ Value of Data to Programs</li> </ul>

These activities yielded several instruments for further consideration and analysis:

<p><b>18</b></p> <p>Classroom Observation Rubrics</p>	<p><b>19</b></p> <p>Dispositional Measures</p>	<p><b>17</b></p> <p>Graduate Surveys</p>	<p><b>12</b></p> <p>Employer Surveys</p>
---	--	--	--

**Stage II** involved working with relevant stakeholders to identify design priorities and criteria for selecting instruments. We administered an online survey to 42 Deans for Impact stakeholders, including member deans, designated data leads, and leading teacher educators. Of those surveyed, 81% responded, and the group reached clear consensus on using a single instrument to collect data per category. The group also reached consensus on three priorities for instrument selection, including:

- 1) ease of implementation across diverse contexts (86% selected as top priority);
- 2) demonstrated reliability and validity whenever possible (83% selected as top priority);
- 3) ensuring maximum adoption by member-led institutions (55% selected as top priority).

The group also achieved consensus on a set of design criteria:

## Design Criteria for Data Category Instruments

	Standards	Scope	Priority Content
Observation of Candidate Instructional Skill	Alignment to State & In TASC standards is a priority for <b>MOST</b>	A comprehensive tool is preferred over more targeted instruments	Delivering Instruction Classroom Environment Designing Instruction
Assessment of Candidate Disposition	Alignment to State & In TASC standards is a priority for <b>SOME</b>	Dispositions are preferred over professional behaviors	Self Reflection Growth Mindset Teaching Self-Efficacy Grit
Graduate Survey	Alignment to State & In TASC standards is a priority for <b>SOME</b>	Graduate feedback is a priority	Preparedness in Core Areas Program Strengths/Weaknesses Employment and Retention
Employer Survey	Alignment to State & In TASC standards is a priority for <b>SOME</b>	Employer feedback is a priority	Relative Effectiveness of Programs/Graduates Programs/Graduate Strengths/Weaknesses Hiring Preferences

After applying the design criteria the group narrowed the list of possible instruments from 66 to a final pool of 10.

**Stage III** involved extensive work by a group of program and faculty leaders at Deans for Impact member-led institutions to deliberate and recommend final measures. In mid-April 2017, these leaders from 12 institutions deliberated on the final 10 instruments over two days. They completed intensive pre-work to understand the arguments for each potential instrument, and participated in a series of structured protocols (e.g. text protocols, chalk talks, jigsaws, and affinity maps) to deliberate on each measure, clarify their thinking, and come to consensus around a set of final instrument recommendations. These recommendations were approved by Deans for Impact member deans, greenlighting the planning process of the CIS pilot year.

### Pilot Year

After two years of development, the CIS Network officially launched for the 2017-18 academic year. During this pilot year we collectively gathered data from 12 institutions in 10 states, yielding a cross-institutional data set with roughly 3,500 teacher-candidates, 400 program graduates, and 100 employers of those graduates.<sup>1</sup>

The CIS Network bookended the pilot year with a set of workshops, called Inquiry Institutes, designed to help programs plan their instrument implementation, analyze their results, and formulate action steps from that data. As programs began collecting common data and implementing plans to improve organizational data cultures, an Advisory Panel comprising of a lead from each institution met monthly to compare notes and iron out wrinkles. Meanwhile, the full network convened once more in the winter to align strategy and build capacity to analyze and plan based on the forthcoming data. Central to that capacity was the development of a data dashboard, the Shared Inquiry Tool, which allows Network members to interrogate their CIS data in real time. The dashboard debuted during the culminating Inquiry Institute in August 2018.

<sup>1</sup> For a full list of participating institutions, refer to Appendix A.

# APPENDIX C

## OVERVIEW OF THE COMMON INDICATORS

---

### CLASS

The Classroom Assessment Scoring System (CLASS)<sup>2</sup> is an observation tool designed to assess the quality of teacher-student interactions associated with positive developmental and academic outcomes for students. CLASS was developed in 2008 by education researchers at the University of Virginia's Center for Advanced Studies in Teaching and Learning (CASTL) and is now administered by Teachstone, which provides training and other services related to CLASS. The CLASS tools used by educator-preparation programs as part of the Common Indicators System are the Upper Elementary and Secondary CLASS, which are comprised of twelve dimensions across three domains. Observers score each dimension on a 1-7 scale defined by detailed descriptors, based on evidence collected during their classroom observation. A typical observation involves one to four cycles of 30 minutes where the observer observes the classroom for 20 minutes and records their scores for 10 minutes. Observers undergo intensive training and must pass an annual exam to demonstrate their ability to reliably score using the CLASS.

### Teacher Beliefs and Mindsets Survey (TBMS)

The Teaching Beliefs and Mindset Survey combines three existing measures:

1. **Short Teachers' Sense of Efficacy Scale**,<sup>3</sup> which assesses the extent to which teachers believe they can influence student engagement, instructional practice, and classroom management. Respondents rate themselves from 1 (nothing) to 9 (a great deal) on 12 statements, like "How much can you use a variety of assessment strategies?"
2. **Short Grit Scale**,<sup>4</sup> which assesses an individual's tendency to persist towards long-term goals. Using a 1-5 scale, respondents rate whether a series of eight statements – like "Setbacks don't discourage me" – are typical of them.
3. **Culturally Responsive Teaching Self-Efficacy Scale**,<sup>5</sup> which assesses how confident teacher-candidates are in their abilities to enact culturally responsive teaching practices. Candidates record a number from 0 (no confidence at all) to 100 (completely confident) in response to statements like "Establish positive home-school relations." A sample of 26 items from this scale are included in the TBMS.

In the CIS Network, the TBMS is administered online to teacher-candidates at the start of their preparation program and then again at the start and end of their clinical experiences.

### Beginning Teacher Survey (BTS)

The Beginning Teacher Survey<sup>6</sup> is based on the New Teacher Preparation Survey (NTPS), developed by researchers at the University of North Carolina's Education Policy Initiative at Carolina for use with graduates of the system's teacher-education programs. The NTPS captures graduates' perceptions of their preparation

---

<sup>2</sup> For the research base underpinning CLASS, see <https://curry.virginia.edu/classroom-assessment-scoring-system>

<sup>3</sup> Tschannen-Moran, M. & Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783-805.

<sup>4</sup> Duckworth, A. L. & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT-S). *Journal of Personality Assessment*, 91(2), 166-174.

<sup>5</sup> Siwatu, K. O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education*, 23(7), 1086-1101.

<sup>6</sup> Bastian, K.C., Sun, M., Lynn, H. (2017). What do graduate surveys tell us about teacher preparation quality? *Education Policy Initiative at Carolina*. Chapel Hill, NC.

experience in five areas: academic background and teaching preparation, teacher preparation quality, teacher preparation program components, current teaching practices, and job satisfaction. CIS Network stakeholders modified the NTPS for use in the CIS by adding an introduction to explain the survey's purpose, revising language to ensure the survey could be implemented successfully across diverse contexts, and eliminating some questions to focus the survey on key areas and reduce the burden on respondents. The resulting Beginning Teacher Survey captures graduates' perceptions of their preparation experience across the five core areas using 26-36 items depending on the graduates' preparation pathway. CIS Network members administer the survey to program graduates in early spring during their first year of full-time classroom teaching.

## Employer Survey

The Employer Survey is a slightly modified version of the Massachusetts Hiring Principal Survey.<sup>7</sup> The survey was developed by the Massachusetts Department of Elementary and Secondary Education, where it is administered annually to all principals who hired a teacher-candidate. For the CIS Network, stakeholders made slight changes to language and survey administration logic, and removed the Massachusetts-specific questions. The resulting survey has seven items on which employers are asked to reflect on the quality of the program graduate, such as "Relative to all other teachers (both novice and experienced) you've worked with, please indicate the extent to which this teacher's performance is significantly above or below average." In the CIS Network, the survey is administered, at a minimum, to employers of recent program graduates who themselves received the Beginning Teacher Survey.

---

<sup>7</sup> Massachusetts Department of Elementary and Secondary Education. (2017). *Educator preparation surveys: Technical report*. Massachusetts Department of Elementary and Secondary Education: Malden, MA.