AIR®

AMERICAN INSTITUTES FOR RESEARCH®

OCTOBER 2016

Aligning Evaluation

How Much Do Teacher Evaluation Rubrics Emphasize Common Core Instruction?

Aligning Evaluation

How Much Do Teacher Evaluation Rubrics Emphasize Common Core Instruction?

OCTOBER 2016

Matthew J. Welch, PhD
Sokoni Davis, EdD
Ruth G. Isaia, PhD
Will T. Johnston
Laura B. Stein
Hannah Jenuwine
Kellie Macdonald

The authors appreciate the collaboration of the following content experts: Mike Garet, Toni Smith, and Beth Ratway.

Contents

Executive Summary

Two efforts that have impacted teachers across the country in recent years are state-mandated teacher evaluation and support systems and statewide adoption of new, more rigorous and focused student learning standards based on the Common Core State Standards. This project explores the level of alignment that exists between these two policies. In particular, we examine the relationship between teacher evaluation rubrics and teaching practices associated with Common Core–aligned standards, including whether alignment between the two is influenced by the rubrics' level of subject-specific content or the state's level of control. Finally, we examine which Common Core–aligned instructional practices are prevalent in the field and which are emphasized less often.

This project, undertaken by American Institutes for Research (AIR) with support from the Bill & Melinda Gates Foundation, contributes to the discussion of whether teacher evaluation rubrics contain specific instructional guidance to teachers that will help them teach the things they have been asked to teach (Common Core State Standards) in the manner they have been asked to teach them.

To conduct this assessment, the research team created the AIR Alignment Protocol (AIR Protocol), a list of instructional practices associated with Common Core–aligned instruction. The team identified eight common instructional practices that observers of Common Core–aligned instruction would expect to see in both English language arts and literacy (ELA-L) and mathematics classrooms, as well as four practices specific to mathematics and three specific to ELA-L.

We then applied the AIR Protocol to 45 teacher evaluation rubrics from nonprofit organizations, charter management organizations, public school districts, and state model rubrics. Of this total, 37 were general instruments, or those meant to apply to lessons of any subject in Grades K–12. Eight were subject specific, meant to assess ELA-L instruction or mathematics instruction. We coded overall alignment for all 45 instruments, calculating alignment scores for all instruments between 0 and 1.

Overall, alignment scores were low but varied widely. The mean alignment score for all 37 general rubrics scored was .23, with a range of 0 to .75, although 80% scored at least .25. Rubrics with particularly strong alignment scores were Denver's LEAP (.75), charter operator Alliance College-Ready (.54), nonprofit TNTP (.50), The KIPP Network (.42), and state model rubrics from Tennessee (.44) and Colorado (.42). The strengths of these instruments include a focus on instruction—rather than other professional attributes of teachers—that included ample details meant to give teachers feedback. Rubrics with higher alignment scores also often included specifically labeled Common Core instructional shifts or broad, general indicators with subject-specific subindicators.

The eight subject-specific rubrics were notable for their depth of detail in several instructional practices that were less common, or less well defined, in other instruments. Practices such as student-to-student discourse, challenging teacher questioning, use of evidence, and integrating disciplines within an academic subject were generally more likely to be present and well defined in nonprofit organizations' subject-specific instruments. On average, subject-specific rubrics (from nonprofit organizations) received higher alignment scores than non-subject-specific rubrics.

The team explored how the level of flexibility afforded to an organization using the rubric might impact alignment. In general, varying levels of state control did not seem to influence local alignment as much as a strong state model. The highest scoring general rubric came from a state with a moderate level of state control and a strong state model (Colorado), and some district-level rubrics from states with low levels of state control also earned relatively strong alignment scores. Charter operators, with significant flexibility to design their own instruments, generally scored well: Four out of five rubrics from charter organizations received higher alignment scores than half of all rubrics scored.

Of the eight common practices we identified, some were commonly emphasized in the sample of rubrics and others were underemphasized. Most rubrics emphasize teachers (a) aligning lessons to standards, (b) using higher order questioning, and (c) assessing understanding. Most rubrics also make some effort to stress cognitively demanding instruction, although this practice was not clearly defined in many cases. Only a handful of rubrics stressed teachers (a) using technology in their lessons, (b) incorporating academic vocabulary, and (c) integrating different concepts within a subject. Having students use evidence to support answers and viewpoints was underrepresented and poorly defined across the sample.

In light of these findings, three steps are recommended in the field of teacher evaluation, particularly in the development of teacher evaluation rubrics:

- Increase emphasis on instruction in general in teacher rubrics, and in particular in underrepresented instructional practices.
- Include additional details on desired practices already included in teacher rubrics, allowing instruments to provide teachers with more specific feedback on their practice.
- Add additional subject-specific content, such as specific indicators or even separate instruments, to allow for unique and in-depth feedback for instruction in different subject areas.

Introduction

Teaching is a complex profession. Educators manage fluid learning environments, where they must reconcile student needs and external expectations by making innumerable decisions and adjustments, often in the middle of lessons. Although teachers certainly influence their own classroom practice, so do other factors, including educational policies made at the local, state, and federal levels. Educational policy initiatives are generally well-intentioned efforts to improve teaching and learning (or their antecedents), but these initiatives also are often envisioned and implemented independently of one another. Teaching becomes increasingly complex when local educators face the dually difficult tasks of not only reconciling the demands of each policy to their local context, but also in trying to weave together different policies that were not coordinated with one another.

Two efforts that have impacted teachers across the country in recent years are state-mandated teacher evaluation and support systems and the statewide adoption of new, more rigorous and focused student learning standards based on the Common Core State Standards. This project explores the level of alignment that exists between these two policies, suggesting to what degree their level of overlap or contrast might serve as a unified, cohesive attempt to improve teaching and learning or as two independent forces that pull teachers in competing directions and make classrooms more complex than they already are.

In particular, we explore the relationship between teacher evaluation rubrics and teaching practices associated with Common Core–aligned standards, including whether alignment between the two is influenced by the rubrics' level of subject-specific content or the state's level of control. Finally, we examine which Common Core–aligned instructional practices are prevalent in the field and which are emphasized less often.

Like all improvement efforts that are concurrently implemented, it is imperative that evaluation and the teaching of new standards be "guided by a common framework" if increased student learning is to result (Newmann Smith, Allensworth, & Bryk, 2001, p. 297).

In the logic of systemic reform, all other parts of the education system that are designed to help guide teachers' classroom instruction—curriculum, assessments, preservice educator preparation, in-service professional learning, educator performance evaluation, and school organization—would be aligned to these student learning standards. Statewide student learning standards detail what students should know and be able to do. These learning standards are an expression of what the state polity believes the education system should work toward helping students achieve. They generally state or imply certain desirable teaching practices that educators might employ as well. If proper alignment was achieved, student learning standards would themselves lend coherence within and among the other parts of the system (see Smith & O'Day, 1991). Such coherence, as many argue, is necessary for improved student learning among all students (e.g., Cohen, 2010/2011; Darling-Hammond, 2013) so that teachers are clear on what students should be learning and when.

One piece of this system, the teacher evaluation rubric, contains both explicit instructional guidance and strong incentives to follow such guidance. Polikoff and Porter (2014) suggest, however, that some current measures of teacher effectiveness may not be appropriately aligned to other teacher demands.

In principle, if the guidance from evaluation tools were well aligned with, and in support of, student learning standards, it could "assist [teachers] in identifying where changes [in practice] are needed" (Porter, 1995, p. 24). However, using teacher evaluation as a tool for accountability and improvement effectively requires explicit connections between statements about desired practices and the tools to evaluate those practices so that teachers, evaluators, and other key stakeholders receive consistent messages about what is expected in instructional practice (Hayes & Lillenstein, 2015). Teachers need clear and consistent feedback, and instruments based on contrasting frameworks may incentivize instructional changes that are not effectively aligned with another key part of the system, student learning assessments (Firestone, 2014). Many states and districts describe teacher evaluation systems as promoting and supporting teachers' professional growth, but few have worked to intentionally and systematically link evaluation and professional development systems in practical ways (Smylie, 2014).

Creating coherence between concurrent reforms, and thereby improving instruction at scale, has always been difficult (Elmore, 2005). Some hoped new sets of learning standards would help ensure consistency and rigor (Smith & O'Day, 1991). However, many teacher evaluation systems are uniform, and not specific to grade or subject, without specifics on instruction. Previous work by American Institutes for Research (AIR) found that "Teacher evaluation reforms are happening in parallel with, but often disconnected from, Common Core implementation. Of the 46 states and the District Columbia that adopted and are implementing the Common Core, at least 36 states and the District of Columbia also now require the use of multiple measures in their teacher evaluation systems" (Leo & Cogshall, 2013, p. 4). These systems are growing in their widespread adoption and in their individual complexity, and work is needed that examines how closely related they are.

This project, undertaken by AIR with the support of the Bill & Melinda Gates Foundation (Gates Foundation), focuses on a key aspect of many teacher evaluation systems: the teacher evaluation rubric. In particular, the research team sought to assess the level of alignment between teacher evaluation rubrics and teaching practices associated with the Common Core, such as increased rigor in classroom instruction. Our purpose was to learn if the messages teachers are receiving from these evaluation tools line up with the explicit and implied demands of the new student learning standards most states have adopted.

These rubrics, often instruments with at least some (if not all) of their components devoted to observing instruction, can be powerful tools to support teachers and evaluate their instruction (see Taylor & Tyler, 2012). Several observers of observation-based teacher evaluation systems have studied methods for their use (Cantrell & Kane, 2013; Curtis, 2012; Donaldson et al., 2014), the need for observer training (Whitehurst, Chingos, & Lindquist, 2014), and concerns about methodological reliability (Milanowski et al., 2014). Little work has focused on the substance of instruction (Hill & Grossman, 2013). We aim to close this gap by assessing whether teacher evaluation rubrics contain specific instructional guidance to teachers that will help them teach the things they have been asked to teach (the Common Core) in the manner they have been asked to teach them.

To conduct this assessment, the research team began by identifying eight common instructional practices that observers of Common Core–aligned instruction would expect to see in both English

language arts and literacy (ELA-L) and mathematics classrooms.[1] We then developed a set of constituent elements, or specific instructional moves, that one might expect to see any classroom, and then more specific moves that might be observed in ELA-L and mathematics lessons, specifically. Our team also developed a list of instructional practices and constituent elements particular to ELA-L and mathematics teaching. Finally, we applied the resulting protocol to a set of teacher evaluation rubrics, selected by the Gates Foundation, to assess their alignment to our list of Common Core–associated teaching practices.

Overall, our findings reveal a lack of detailed attention in most rubrics to the kinds of specific, rigorous instructional practices that Common Core–aligned instruction would be expected to include. Although the rubrics evaluated were a convenient, not a representative sample, of the 45 rubrics evaluated, they included state-level model rubrics, as well as instruments from local public school districts, nonprofit organizations, and charter school operators. Overall:

- 0% showed even partial alignment to all of the instructional practices;
- Just one rubric showed alignment to a majority of the practices (16 of 18 general instructional elements);
- The three most commonly observed practices (stressing alignment to state standards, assessing understanding, and engaging students with cognitively demanding tasks) were aligned to approximately 50% of rubrics; and
- Six rubrics showed zero alignment to any of the practices; another four showed alignment to only one instructional element. A handful of rubrics contained too little content to score. For example, three Kentucky districts appear to have adopted the Kentucky state model but provided shorter documents that lack the detail of the state model and are likely walk-through checklists. Two other Kentucky districts, Clay and Warren, provided documents with too little content to consider.

We further examined these trends in terms of the context in which the rubrics are used, possible variations between those from public school districts and those from other sources, and the representation of the AIR-identified instructional practices in the field as a whole. Highlights from these findings, detailed later in the report, include:

- On average, stronger alignment was found in rubrics from charter schools, nonprofits, and a selection of strong state model rubrics, especially when compared with most district instruments.
- Although there is no clear relationship between level of state control and alignment, a strong state model coupled with at least moderate levels of state control appears to result in stronger alignment.
- Eight of the nine rubrics scored were subject specific, and four of nine were specific to a range of grade levels. On average, subject-specific rubrics (from nonprofit organizations) received higher alignment scores than non-subject-specific rubrics.

[1] For more information on the development of the AIR Alignment Protocol, see Appendix D.

Methods and Sample

This report is the last in a series of projects and reports conducted by AIR with support from the Gates Foundation, assessing alignment between teacher rubrics and Common Core–aligned teaching practices. The AIR Alignment Protocol (AIR Protocol) was developed in spring 2015, listing essential Common Core–aligned practices in ELA-L and mathematics (the documents that served as the basis for the AIR Protocol are highlighted in Appendix D). This report builds on others that have been presented to the Gates Foundation in the past two years, including the first report to assess alignment between Common Core–aligned practices and the teacher evaluation rubrics of nine states (see Welch, Freed, et al., 2015), and the second report to assess alignment of five rubrics from four nonprofit organizations as well as teacher evaluation rubrics from 17 public school districts in two states (see Welch, Potemski, et al., 2015).[2]

This final report takes a cumulative look at AIR's work in this area, including results from the fall 2015 analysis and a rescoring of several states from the spring 2015 analysis. We collected or attempted to collect 101 instruments in total, all at the request of the Gates Foundation, from states, district, charter operators, and nonprofit organizations. Of these, 45 rubrics were scored. Of the 56 that were not able to be scored, 17 contained too little content to be considered, even when they were likely facsimiles of state models. These shorter instruments were often walk-through checklists or short forms. The other 39 rubrics directly duplicated their state model rubrics, meaning those scores would be identical to other instruments scored as part of this process.

Included in this report are alignment scores on 45 rubrics from several sources:

- Nine rubrics from four nonprofit organizations.
- Teacher evaluation rubrics from 25 public school districts in five states—Colorado, Florida, Kentucky, Massachusetts, and Tennessee.
- State-level model rubrics from six states—California, Colorado, Kentucky, Massachusetts, North Carolina, and Tennessee.
- Rubrics from five charter school networks.

Thirty-seven of these instruments were general, universal rubrics, designed for use in any Grades K–12 classroom or lesson; eight were subject-specific, focusing on lessons and instructional practices particular to ELA-L or mathematics. Some general rubrics contained specific indicators for particular subjects or grades. Four of the subject-specific rubrics were focused on particular grade ranges. All were scored between September 2015 and July 2016.

The first group of five rubrics scored were published by four organizations: TNTP, the Southern Regional Education Board (SREB), Charlotte Danielson's reorganized Framework for Teaching (FFT), and Student Achievement Partners (SAP). SAP divides its teacher evaluation instruments for upper and lower grade levels. Some organizations publish separate rubrics for ELA-L and mathematics.

[2] See Welch, Potemski, et al.'s (2015) report for details about revisions to the AIR Protocol as well as reliability statistics and procedures.

The nonprofit organizational rubrics scored were as follows:

- **TNTP:** One rubric, common to ELA-L and mathematics
- **SREB:** Two rubrics, one for ELA-L and one for mathematics
- **FFT:** Two rubrics, one for ELA-L and one for mathematics
- **SAP, Upper Level:** Two rubrics, one for ELA-L (Grades 3–12) and one for mathematics (Grades 9–12)
- **SAP, Lower Level:** Two rubrics, one for ELA-L (Grades K–2) and one for mathematics (Grades K–8)

The state and district rubrics scored were all general instruments from California (one, the state model), Colorado (the state model and six districts), Florida (three districts), Kentucky (the state model and 13 districts), Massachusetts (the state model, one district rubric, and one regional rubric), North Carolina (the state model), Tennessee (the state model and one district), and rubrics from five charter management organizations.

Scoring Protocol and Scoring Procedure

The team scored all 45 teacher evaluation rubrics from the districts, states, nonprofits, and charter networks using the same AIR Protocol. The AIR Protocol consists of three sections—one general, one ELA-L specific, and one mathematics specific:

- Eight instructional practices—the common practices—supporting the adoption of the Common Core–aligned standards in both ELA-L and mathematics classrooms;
- Three instructional practices associated with the adoption of only the ELA-L standards; and
- Four instructional practices supporting the adoption of only the mathematics standards.

The full AIR Alignment Protocol is available in Appendix A.

Each instructional practice consists of a number of constituent elements, or specific instructional moves that make up each practice. The elements define the practice in action. These elements come in two forms, general (describing any lesson in any subject) and subject specific (describing a lesson in either ELA-L or mathematics). Table 1 includes the eight common practices, as well as the number of general elements, mathematics elements, and ELA-L elements for each.

Table 1. List of Common Practices and Number of Elements From AIR Protocol

Common Practice (CP)	Number of General Elements	Number of Mathematics Elements	Number of ELA-L Elements
CP 1: Establish appropriate unit learning goals.	2	3	3
CP 2: Engage students with cognitively demanding tasks.	3	4	4
CP 3: Promote strategic and appropriate use of domain-specific tools and resources.	2	3	3
CP 4: Facilitate student discourse.	3	3	3
CP 5: Engage in purposeful questioning.	3	3	3
CP 6: Integrate key components of content and dispositions of the domain.	1	2	1
CP 7: Promote the use of academic and domain-specific language and vocabulary.	3	3	4
CP 8: Assess important skills and understandings.	1	2	2

In addition to the section containing the eight common practices, the AIR Protocol also contains two sections focused on subject-specific practices and their respective, constituent elements: a section containing four mathematics-specific practices as well as a section containing three ELA-L–specific practices. The mathematics-specific and ELA-L–specific practices are listed in Tables 2 and 3, respectively:

Table 2. List of Mathematics-Specific Practices and Number of Elements From AIR Protocol

Mathematics-Specific Practice	Number of Mathematics Elements
Math 1: Embed mathematics in real-world contexts and emphasize modeling.	2
Math 2: Support students' productive struggle in mathematics.	2
Math 3: Promote the use of multiple representations and connections among them.	2
Math 4: Encourage abstract reasoning.	3

Table 3. List of ELA-L–Specific Practices and Number of Elements From AIR Protocol

ELA-L–Specific Practices	Number of ELA-L Elements
ELA-L 1: Support early reading, foundational skills (Grades K–5 teachers only).	1
ELA-L 2: Teach writing for research, argumentation, and narrative.	4
ELA-L 3: Build independence for all students: Providing a staircase to the College and Career Readiness anchor standards.	4

Aspects of Rubrics Coded

The 45 rubrics varied in several areas relevant to the coding procedure, the resulting alignment scores, and our subsequent analysis.

The 45 scored rubrics generally fell into one of two categories of independence and one of two categories regarding target subjects. These classifications are displayed in Table 4. As shown in the table, all 31 state and district teacher evaluation rubrics were general, or not designed for any particular subject area and for use in any Grades K–12 classroom. An additional six rubrics designed by one nonprofit and five charter operators also were general instruments. Eight of the 45 were subject-specific instruments.

Table 4. Rubric Classification Categories

	General (*N*)	Subject Specific (*N*)
Public Oversight	State and district instruments (31)	None
Independent	TNTP (1) Charter Schools (5)	SAP (4) SREB (2) FFT (2)

Most rubrics were general, although some had clearly labeled subject-specific indicators. Colorado's state model, for example, had several general practices with subject-specific subindicators.

Rubrics varied in other ways, including who created them and could dictate their use, their guiding framework, and length, as well as level of subject-specific content. In Appendix B, Exhibits B1 and B2 illustrate the number of individual indicators that were considered when scoring each instrument. In addition, Exhibit B1 highlights which district rubrics had no subject-specific content to consider when scoring. The exhibit also notes the political context in which each instrument is used, in the form of the level of state control exercised by the respective states regarding the design and use of teacher evaluation rubrics in general. Danielson's FFT was the guiding framework used by several districts and states.

Exhibit B2 in Appendix B highlights that all of the rubrics published by nonprofit organizations, except the one published by TNTP, had separate instruments specific to ELA-L and mathematics. Further, the four published by SAP included separate rubrics for not only ELA-L and mathematics but also students' age and grade ranges.

Coding the Rubrics

Using the AIR Protocol, each of the 37 general teacher evaluation rubrics was coded for alignment by one expert in the ELA-L standards and one expert in the mathematics standards. Each of the eight subject-specific rubrics were coded by two content experts from the same content area.

Rubrics were scored using matrix coding, where each indicator and descriptor of each teacher evaluation rubric is compared with each element of the AIR Protocol. Coding is binary, marking an element as present (1) or not (0). Possible scores on each of the eight common practices, as well as overall alignment scores, range from a low score of 0 to a high score of 1.

Table 5 displays example language from one rubric. The table illustrates that the rubric contains standards that are general categories of professional duties, indicators that are particular professional practices within standards, and descriptors that are specific actions that are components of the indicators. It is the specific instructional actions, such as those in the right-hand columns of Table 5, that coders compared with the elements of the AIR Protocol.

Table 5. Examples of Rubric Language From One District

Quality Standard	Indicator	Descriptor
Fayette, Kentucky (based on Danielson's FFT)		
Planning and Preparation	Designing Student Assessments	■ Teacher's plan for student assessment is aligned with the instructional outcomes; assessment methodologies may have been adapted for groups of students. ■ Assessment criteria and standards are clear. Teacher has a well-developed strategy for using formative assessment and has designed particular approaches. ■ Teacher intends to use assessment results to plan for future instruction of groups of students. ■ All the learning outcomes have a method for assessment. ■ Assessment types match learning expectations. ■ Plans indicate modified assessments for some students as needed. ■ Assessment criteria are clearly written. ■ Plans include formative assessments to use during instruction. ■ Lesson plans indicate possible adjustments based on formative assessment data.

Rubrics that were subject specific were coded slightly differently than the 37 general rubrics. As shown in the AIR Protocol in Appendix A, each of the eight common practices has general instructional elements as well as subject-specific elements for both ELA-L and mathematics. For example, Common Practice 5: Engage in Purposeful Questioning has three general elements that define the practice. In this practice, the teacher uses questioning strategies to:

- Advance student reasoning and understanding (e.g., through higher order questioning);
- Encourage students to explain their thinking; and
- Encourage or require students to justify their claims with evidence.

Rubrics were coded against the general elements if the instrument was primarily a general, or universal rubric. Thirty-seven rubrics were classified as general.

Rubrics were coded against the subject-specific elements if the document was deliberately subject specific. The same instructional practice, Common Practice 5: Engage in Purposeful Questioning,

has three ELA-L-specific elements that define that practice in ELA-L classrooms. In ELA-L lessons, the teacher uses questioning strategies to:

- Require students to read closely to advance their understanding of texts;
- Encourage students to explain their interpretations of understanding of texts; and
- Require students to use evidence from the text to justify increasingly complex analyses.

Rubrics were coded against these subject-specific elements if the instrument was a subject-specific rubric. Eight rubrics were classified as subject specific.

General rubrics with subject-specific indicators could be coded against subject-specific elements for those subject indicators. An indicator needed to have an explicit subject focus to be coded as having any of these practices and elements.

All rubrics were coded against not only the elements of the eight common practices but also the seven subject-specific practices in the AIR Protocol (four in mathematics and three in ELA-L, described in Tables 2 and 3). However, because a rubric or indicator needed to have a deliberate subject-specific focus to receive a possible alignment score, no general rubrics were found to highlight any of the seven subject-specific practices or their constituent elements.

As a result of the AIR Protocol describing the desired practices in both general and subject-specific ways, it is possible to calculate more than one score for general rubrics, one using only the general elements of the eight common practices and one using all available instructional practices in the AIR Protocol. Appendix C contains more information on methods, interrater reliability, and sensitivity measures for each instructional element, as measured using Cohen's kappa. Subject-specific rubrics received only one overall alignment score.

Results

This section presents scores for all of the teacher evaluation rubrics that were assessed. It includes a discussion of findings relative to the rubric type and content, state teacher evaluation policy, and trends in the field of instructional practices, both well represented and underrepresented. Namely, we explore the following overarching questions about alignment:

1. What is the relationship, if any, between rubric type (general or subject specific) and degree of alignment to the Common Core–aligned practices?
2. What is the relationship, if any, between level of state control over teacher evaluation policy and degree of alignment to Common Core–aligned practices?
3. What are the most and least common instructional practices, supportive of the Common Core State Standards, represented in teacher evaluation rubrics?

In general, overall alignment of state, district, and other rubrics scored to the elements and practices of the AIR Protocol was relatively low. One of the reasons for the low-to-moderate rates of alignment

is lack of explicit details in the observation rubrics, ambiguity in phrasing that meant that specific teacher behaviors in the AIR Protocol could not be clearly assessed, or a lack of instructionally focused content in some instruments. Some rubrics were excluded from analysis or received overall scores of 0 based on their brevity.

As discussed in the preceding section (and detailed in Appendix C), most rubrics received two different scores, depending on the rubric's level of subject-specific content. Table 6 presents both the scores—general and subject specific—for all 37 general rubrics as well as total alignment scores for the eight subject-specific rubrics. "Average Alignment Score" refers to the average alignment score for the eight common practices, using general elements only for general rubrics and subject-specific elements of the eight common practices and the seven subject-specific practices for subject-specific rubrics. State models are indicated in bold, and the less applicable score is grayed out. The general elements mean for all rubrics scored was .23, with a range of 0 to .75 alignment.

For all state models and district rubrics scored, Table 6 also shows the level of state control over teacher evaluation policy. These state control categories are explained in more detail in the Level of State Control section.

The instruments are ranked by the most applicable score based on the focus of the rubric. Several trends are evident in this ranking of rubrics:

- The teacher evaluation rubric with the highest general score, Denver's LEAP, is one that includes two thirds of its indicators focused on instructional practices and deliberately labeled Common Core–aligned instructional shifts but contains no indicators grounded in specific academic subjects.
- The state model rubric from Colorado is the only instrument in the top quartile in both general and subject-specific scores. The instrument contains several broadly defined instructional practices as well as subject-specific subindicators.
- Rubrics in the bottom quartile of the table, particularly those receiving alignment scores of 0, contain little detail for either evaluating teachers or offering feedback.
- Most general rubrics see their alignment scores fall when accounting for the additional, subject-specific instructional practices on the AIR Protocol. One exception is Silverton, Colorado, which is one of the rubrics with few indicators and no defining detail for any of those indicators. However, a small number of these indicators are subject specific, leading to some subject-specific alignment and higher subject-specific score then a general alignment score.

Table 6. Summary of Alignment Scores by Rubric, Average of Eight Common Practices

Quartile	State/ Charter (CH)/ Nonprofit Organization (NP)	Level of State Control	District/Charter/ Organization	Average Alignment Score (General Elements Only)[a]	Average Alignment Score (General and Subject-Specific Elements)
Top 25%	CO	Moderate	LEAP (Denver)	0.75	0.19
	CH	n/a	Alliance College-Ready	0.54	0.14
	NP	n/a	TNTP	0.50	0.13
	NP	n/a	SAP Lower ELA-L	–	0.45
	TN	**High**	**State Model (TEAM)**	**0.44**	**0.11**
	NP	n/a	Danielson's FFT Clusters-ELA-L	–	0.43
	CH	n/a	KIPP	0.42	0.15
	CO	**Moderate**	**State Model (CO State Model)**	**0.42**	**0.34**
	NP	n/a	SAP Upper Math	–	0.42
	NP	n/a	SAP Lower Math	–	0.38
	FL	Low	Pinellas	0.38	0.09
	TN	High	Memphis	0.38	0.09
Top 50%	NP	n/a	SAP Upper ELA-L	–	0.36
	KY	Low	Fayette	0.35	0.09
	KY	**Low**	**State Model (KY Framework)**	**0.35**	**0.09**
	NP	n/a	SREB ELA-L	–	0.34
	KY	Low	Floyd	0.31	0.08
	CA	**Low**	**CA CSTP (CA State Model)**	**0.29**	**0.07**
	CH	n/a	YES Prep	0.29	0.07
	FL	Low	Broward	0.29	0.07
	NP	n/a	SREB Mathematics	–	0.29
	NC	**High**	**State Model (NC)**	**0.29**	**0.07**
	CH	n/a	Achievement First	0.27	0.07
Bottom 50%	NP	n/a	Danielson's FFT Clusters-Math	–	0.25
	KY	Low	McLean	0.23	0.06
	FL	Low	Brevard	0.21	0.05
	KY	Low	Nelson	0.19	0.05
	MA	Moderate	CREST (multiple districts)	0.19	0.05
	MA	**Moderate**	**State Model (MA Framework)**	**0.19**	**0.05**

Quartile	State/ Charter (CH)/ Nonprofit Organization (NP)	Level of State Control	District/Charter/ Organization	Average Alignment Score (General Elements Only)[a]	Average Alignment Score (General and Subject-Specific Elements)
	CO	Moderate	Durango	0.17	0.04
	KY	Low	Pike	0.17	0.04
	KY	Low	Russell County	0.17	0.04
	KY	Low	Wayne	0.17	0.04
Bottom 25%	CO	Moderate	Thompson	0.10	0.03
	KY	Low	Henderson	0.10	0.03
	MA	Moderate	Easton	0.06	0.02
	CO	Moderate	Jefferson County	0.04	0.01
	CO	Moderate	Silverton	0.04	0.12
	KY	Low	Trimble	0.04	0.01
	CH	n/a	STRIVE	0	0
	CO	Moderate	Garfield	0	0
	KY	Low	Clinton	0	0
	KY	Low	Mercer	0	0
	KY	Low	Boyd	0	0
	KY	Low	Russell Independent	0	0

[a] Throughout the report, "Average Alignment Score" refers to the average alignment score for common practices, general elements only, unless otherwise specified.

State Models, Charter, and Nonprofit Approaches

Key Findings:

- State model rubrics averaged .33. Eighty percent (five of six) of state model rubrics received alignment scores of at least .25, whereas only 24% of district rubrics received alignment scores of .25 or greater, indicating model rubrics are not always followed with fidelity.
- Four out of five rubrics from charter organizations received higher alignment scores than half of all rubrics scored. The other rubric received an alignment score of 0.
- On average, rubrics from nonprofit organizations received the highest overall alignment scores, as compared with other rubric types scored, including alignment to both general and subject-specific elements for subject-specific rubrics, with all rubrics receiving scores of .25 or higher.
- All but one rubric from nonprofit organizations was subject-specific, resulting in higher total alignment scores that included subject-specific elements. In contrast, none of the state, district, or charter organization rubrics scored were subject-specific and therefore received lower total alignment scores (general elements only) on average.

Although the majority of rubrics scored came from individual districts, AIR also scored six state models, rubrics from five charter organizations, and nine rubrics from nonprofit organizations. Table 7 shows the average alignment score for each rubric type, along with the percentage of rubrics showing various levels of alignment.

Table 7. Average Rubric Score, by Type

Type	*N*	Average Alignment Score	% Rubrics Scoring > .50	% Rubrics Scoring Between .25–.49	% Rubrics Scoring < .25
Districts	25	.17	4%	20%	76%
Charters	5	.30	20%	60%	20%
State models	6	.33	0%	83%	17%
Nonprofit—General	1	.50	100%	0%	0%
Nonprofit—Subject specific[a]	8	.37	0%	100%	0%

[a] Alignment scores for subject-specific rubrics include alignment to subject-specific and general elements.

Eight of the rubrics from nonprofit organizations were subject specific (four mathematics and four ELA-L) and received average alignment scores that include general and subject-specific elements (as described in the Scoring section). The other rubric, from TNTP, was not subject specific. As a result, the TNTP rubric received an alignment score that considers alignment to general elements only and is included with the charter rubrics scored for the analysis by rubric type described in this section.

On average, state models and rubrics from charter organizations and nonprofit organizations (both general and subject specific) showed higher levels of alignment than individual district rubrics.

State Models

The majority of state models received alignment scores of at least .25. In contrast, the majority of district rubrics scored received alignment scores of less than .25. Overall, state models were more aligned to the AIR Protocol than district rubrics, and the lowest scoring state model (Massachusetts) received a higher score (.19) than more than 25% of all rubrics scored.

The state models influence the level of alignment in district rubrics. In cases where state models were coded as aligned to one of the elements in the eight common practices, the majority of the districts in that state also were coded as full or partially aligned in that practice area. The converse also is true. For example, the state model for Kentucky received a score of 0 (nonalignment) in three out of the eight common practices: (1) Establish Appropriate Learning Goals, (6) Integrate Key Components of the Domain, and (7) Promote Use of Academic and Domain-Specific Language. All of the districts in Kentucky also received a score of 0 for nonalignment in the same Common Practice areas. Not all districts are faithful to their state's model, and some states, like Kentucky, grant wide flexibility to local actors. Nonetheless, state models obviously have some influence over districts, and local evaluators benefit from a strong state model.

Charter Organizations

All five of the charter rubrics scored, and one of the nonprofit rubrics scored (TNTP), were considered general rubrics.[3] As such, these six rubrics were analyzed together and findings are presented here, whereas the subject-specific nonprofit rubrics were analyzed separately. Throughout this section, references to the "charter rubrics scored" includes the TNTP rubric. All but one charter rubric received an alignment score of at least .25 on the AIR Protocol, the TNTP rubric scored .50, and one charter rubric scored higher than .50 (Alliance College-Ready, .54).

Of the charter rubrics scored, Alliance College-Ready Charter received the highest average alignment score to the eight common practices and general elements. The strengths of this rubric include detailed expectations for students to think and speak critically at higher instructional levels. The rubric also details the lesson-planning process to indicate higher levels in instructional practices and student outcomes. In addition to these higher order thinking skills, the rubric indicates expectations for students to support their ideas and answers with evidence, which is not indicated in many other rubrics. Three of the other five charter rubrics scored (KIPP, YES Prep, and Achievement First) also received relatively high alignment scores, scoring higher than 50% of all rubrics scored, indicating a greater focus on instructional practice than some other instruments assessed.

STRIVE Charter showed no alignment to any of the eight common practices or general elements. The STRIVE rubric indicated several elements for classroom processes and structures, such as discipline strategies and procedures and classroom climate observations, but the rubric did not focus enough or give explicit details on instructional practices and student outcomes to show alignment to our eight common practices. The instructional practices that are mentioned are vague and lack specific examples of implementation. For example, for Common Practice Area 5 (Engage in Purposeful Questioning), the rubric does mention that students are required to think at higher levels but does not offer any descriptions of these tasks. The rubric mentions that students will "turn and talk" to discuss their answers but does not indicate an expectation for students to provide evidence or reasoning for their thinking.

Alignment to Common Practices

Common Practice 1 (Establish Appropriate Learning Goals) received the highest average score of all eight of the practice areas. All but one charter rubric was aligned to at least one element of this practice; Strive Charter is the only charter rubric scored that showed no alignment to this practice area. Although the STRIVE Charter rubric indicates lesson planning, there is no mention of aligning lessons to state standards or a progression of content over time. Achievement First Charter was the only charter rubric scored that received partial alignment in this practice area. The rubric indicates an alignment of lessons to state standards but does not indicate a progression of content and skills over time. Alliance Charter, KIPP Charter, and YES Prep Charter, along with TNTP, all received scores showing full alignment to both elements by indicating an alignment to state standards and progression over time of content, skills, and materials through lesson plans, units, or curriculum maps.

[3] TNTP, the only general nonprofit rubric scored, is included in these analyses because the rubric was more comparable in approach to the other charter rubrics than to the other nonprofit rubrics.

Common Practice 7 (Promote Use of Academic and Domain-Specific Language and Vocabulary) received relatively low alignment scores across charter rubrics scored, with an average score of .11. Alliance College-Ready and Achievement First received partial alignment for this practice by indicating an expectation for students to use academic vocabulary in the classroom; however, two thirds of charter rubrics scored received a score of 0, indicating no alignment to this practice. Many of the rubrics did not explain ways of modeling vocabulary usage or indicate explicit teaching strategies for assessing student vocabulary mastery during instruction. The rubrics also did not explain teacher expectations of student outcomes. The rubrics also lacked specific ways that students will use vocabulary in oral and written activities, evidence of alignment to Common Practice 7.

None of the charter rubrics scored showed alignment to Common Practice 3 (Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources) or Common Practice 6 (Integrate Key Components of the Domain).

Common Practice 3. With regard to the Common Practice 3 element, *Strategic use of tools or resources, including technology, to support work in the domain,* many rubrics failed to indicate the use of specific technology components during instruction or student-guided activities such as the 21st century learning competencies and skills or multimedia activities and presentations. If technology was mentioned in the rubrics, the task was vague and did not connect to teaching and learning. With regard to the Common Practice 3 element, *Student involvement in a decision-making process through which to determine which tool is appropriate to the work*, many rubrics failed to mention opportunities for student voice in decision making or the teacher as a facilitator in student collaborative practices in choosing appropriate tools to guide learning and creativity.

Common Practice 6. With regard to the Common Practice 6 element, *Attends to and integrates key components, proficiencies, and dispositions within the discipline*, many of the rubrics fail to indicate specific descriptions and examples of components indicating teacher knowledge of content and pedagogy. The rubrics also show no evidence of instruction for addressing the needs of diverse learners or a range of expected student and instructional outcomes.

Nonprofit Organizations

On average, nonprofit rubrics—both general (TNTP) and subject specific—scored higher than any other rubric type. Eight of the nine nonprofit rubrics scored were subject specific (four ELA-L and four mathematics). One set of instruments, from SAP, also included separate rubrics for upper and lower grade levels in both ELA-L and mathematics. The other nonprofit rubric scored, TNTP, was not specific to any subject area. Average alignment scores for subject-specific rubrics consider alignment to subject-specific elements in addition to general elements and generally received higher total alignment scores than general rubrics, whose scores includes alignment to general elements only.

Table 8 shows the average alignment score, general elements only, and the average alignment score that includes both general and subject-specific elements for both types of rubrics scored.

Table 8. Average Alignment Score for General and Subject-Specific Rubrics

Type	Number of Rubrics Scored	Average Alignment Score (General Elements Only)	Average Alignment Score (General Elements and Subject-Specific Practices)
Subject-specific rubrics	8	n/a	.37
General rubrics	37	.23	.07

Note. n/a = not applicable.

Table 9 shows how the rank order of rubrics scored, by average alignment scores changes, when subject-specific elements are included in the overall alignment score, for both general and subject-specific rubrics. The only general model that maintains a relatively high alignment score (.34) is the Colorado state model rubric, which contains several broad instructional indicators and subsequent subindicators that are specific to various subjects and grade levels. The other top rubrics scored are all subject-specific nonprofit rubrics, which each scored at least .25.

Table 9. Average Rubric Score, Ranked by Total Alignment (General and Subject Elements)

State/ Charter (CH)/ Nonprofit Organization (NP)	Rubric Type	District/Charter/Organization	Alignment Score (General Elements and Subject-Specific Indicators)	Alignment Score (General Elements Only)
NP	Subject specific	SAP Lower ELA-L	0.45	–
NP	Subject specific	Danielson's FFT Clusters-ELA-L	0.43	–
NP	Subject specific	SAP Upper Math	0.42	–
NP	Subject specific	SAP Lower Math	0.38	–
NP	Subject specific	SAP Upper ELA-L	0.36	–
CO	General	State Model (CO State Model)	**0.34**	**0.42**
NP	Subject specific	SREB ELA-L	0.34	–
NP	Subject specific	SREB Mathematics	0.29	–
NP	Subject specific	Danielson's FFT Clusters-Math	0.25	–
CO	General	LEAP (Denver)	0.19	0.75
CH	General	KIPP	0.15	0.42
CH	General	Alliance College-Ready	0.14	0.54
NP	General	TNTP	0.13	0.50
CO	General	Silverton	0.12	0.04
TN	General	State Model (TEAM)	**0.11**	**0.44**
FL	General	Pinellas	0.09	0.38

State/ Charter (CH)/ Nonprofit Organization (NP)	Rubric Type	District/Charter/Organization	Alignment Score (General Elements and Subject-Specific Indicators)	Alignment Score (General Elements Only)
TN	General	Memphis	0.09	0.38
KY	General	Fayette	0.09	0.35
KY	General	State Model (KY Framework)	**0.09**	**0.35**
KY	General	Floyd	0.08	0.31
CA	General	CA CSTP (CA State Model)	**0.07**	**0.29**
CH	General	YES Prep	0.07	0.29
FL	General	Broward	0.07	0.29
NC	General	State Model (NC)	**0.07**	**0.29**
CH	General	Achievement First	0.07	0.27
KY	General	McLean	0.06	0.23
FL	General	Brevard	0.05	0.21
KY	General	Nelson	0.05	0.19
MA	General	CREST (multiple districts)	0.05	0.19
MA	General	State Model (MA Framework)	**0.05**	**0.19**
CO	General	Durango	0.04	0.17
KY	General	Pike	0.04	0.17
KY	General	Russell County	0.04	0.17
KY	General	Wayne	0.04	0.17
CO	General	Thompson	0.03	0.10
KY	General	Henderson	0.03	0.10
MA	General	Easton	0.02	0.06
CO	General	Jefferson County	0.01	0.04
KY	General	Trimble	0.01	0.04
CH	General	STRIVE	0	0
CO	General	Garfield	0	0
KY	General	Clinton	0	0
KY	General	Mercer	0	0
KY	General	Boyd	0	0
KY	General	Russell Independent	0	0

Each of the subject-specific rubrics for SAP, Danielson's FFT, and SREB showed at least partial alignment to the subject-specific elements of all eight common practices.

SAP. SAP Upper and Lower Math scored the highest for mathematics subject-specific rubrics, with total alignment scores of .42 and .38, respectively, and at least partial alignment to seven of the eight common practices. The strengths of the SAP mathematics rubrics for upper and lower grades include lessons that are rigorous and aligned to standards of mathematical practice. The rubric also explicitly states the expectation of students to explain their problem-solving processes through higher order questioning and use of academic language. SAP Lower ELA-L received the highest overall score (.45) for ELA-L–specific rubrics scored.

Danielson's FFT. The widely adopted FFT has been reorganized into two subject-specific cluster documents. The reorganized FFT received the second highest score of ELA-L subject-specific rubrics (.43) and showed at least partial alignment to seven of the eight common practice areas. The strengths of this rubric are the explicit details of instructional outcomes for ELA-L, including an emphasis on literacy standards and the expectations for using evidence in classroom discussions. The rubric details teacher modeling of skills and opportunities for student practice in mastery of ELA-L standards. The rubric also states opportunities to assess learning of all students through teacher reflection, formative and summative assessments with course corrections for ELA-L instruction based on the results.

SREB. The strengths of the SREB ELA-L rubric, which received an alignment score of .34, explicitly indicate the expectation of students using textual evidence to support their answers through higher order questioning and activities. The rubric also indicates in detail the use of classroom rubrics for students to assess their work as a form of assessment for ELA-L and mathematics subject areas. The mathematics rubric, which scored .29, indicates details of student discourse by discussing mathematics problems and analyzing student answers to problems. However, SREB showed no alignment to four out of eight common practices for ELA-L or to four out of eight practices in mathematics. The areas that showed nonalignment in ELA-L were the lack of opportunities for students to engage in peer-to-peer and group collaboration as well as no opportunities indicated for student voice. The mathematics rubric failed to mention alignment to mathematics standards or progression of skills and content over time. Neither the mathematics nor the ELA-L SREB rubrics showed alignment to Practice Area 3; the rubrics do not indicate the use of technology or mathematical tools to solve problems or address real-world applications in either subject-specific area.

Alignment to Common Practices

Common Practice 4 (Facilitate Student Discourse) had the highest alignment across subject-specific rubrics scored, with an average of 58%. Many rubrics indicate opportunities for students to interact with peers on tasks through collaboration. The rubrics also indicate teachers engaging students to provide evidence for their responses. Common Practice 3 (Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources) had the lowest alignment across subject-specific rubrics scored, with an average score of 8%. Although many of the rubrics indicate opportunities for student discourse and collaboration, the rubrics do not indicate opportunities for student voice in decision making. The

ELA-L and mathematics rubrics fail to indicate the use of specific technology components during instruction or student-guided activities such as the 21st century learning competencies and skills and multimedia activities and presentations. The SAP Charter rubrics for the upper and lower schools showed partial alignment to this common practice by detailing the strategic use of mathematical tools to solve problems.

See Appendix B for alignment scores by rubric on the eight common practices.

On average, subject-specific rubrics (from nonprofit organizations) received higher alignment scores than non-subject-specific rubrics. A detailed discussion of alignment to subject-specific practices and elements is included in the section Prominent Instructional Practices.

Level of State Control

Key Findings:

- Level of state control, as an isolated variable, does not appear to predict higher or lower levels of alignment.
- The highest-scoring rubric was a district instrument that came from a state with a moderate level of state control and a strong state model. However, other district-level rubrics from states with low levels of state control also earned relatively strong alignment scores.
- Alignment scores varied more widely across district-level rubrics from states with low or moderate levels of state control.

Although traditionally the purview of districts, control over teacher evaluation policy has increasingly shifted to the state level in recent years. Still, wide variation exists in how much control individual states exercise in designing and implementing new evaluation systems. The Center on Great Teachers and Leaders (GTL Center) at AIR has identified and defined three levels of state control—low, moderate, and high—and categorized each state accordingly. Table 10 displays the GTL Center's definitions and criteria for each level.[4]

[4] See http://resource.tqsource.org/stateevaldb/StateRoles.aspx.

Table 10. Forms of State Control in Teacher Evaluation

Low	Moderate	High
District Evaluation System With Required Parameters	**Elective State-Level Evaluation System**	**State-Level Evaluation System**
■ State provides guidance to districts and plays only a small role in design and implementation of the evaluation system at the local level. ■ Guidance is general: Requires LEAs [local education agencies] to include certain components (observations, professional responsibilities, measures of student learning) but allows LEAs wide latitude in selecting or creating those components. ■ Provides screening or approval to ensure compliance with state regulations on frequency, training, grievance procedures, etc.	■ State strictly interprets some aspects of federal and state legislation but allows flexibility in other aspects. ■ State may mandate use of student growth measures, models, and weights while leaving observation, protocols, or additional measures up to LEAs. ■ State may offer a model framework with an observation instrument but allow districts to choose alternative models and instruments if they meet state criteria.	■ State provides strict interpretation of federal and state legislation. ■ State prescribes the requirements for the evaluation model(s). ■ State determines components, measures, frequency, and types of evaluators. ■ All districts must implement the state-designed model with little flexibility.

For this study, AIR used the GTL Center's categorizations to examine the relationship between level of state control and alignment of state model evaluation rubrics and, more importantly, district rubrics. Although our analysis of alignment, by level of state control, shows some interesting patterns, which are described in the findings here, the sample size and composition limit AIR's ability to draw any strong conclusions about the relationship between level of state control and alignment. All scored rubrics were selected by the Gates Foundation from jurisdictions of interest, making the sample uneven in terms of levels of state control. These limitations are acknowledged, where appropriate, throughout.

Relationship Between Level of State Control and Overall Alignment

Some research suggests that local flexibility for districts and schools to implement policies like teacher evaluation may correlate with increased instructional effectiveness and achievement (Honig & Hatch, 2004). Most states currently exercise low or moderate levels of state control of teacher evaluation systems. Although the sample was not evenly distributed across the three categories of control, these findings suggest that significant flexibility will not necessarily lead to significantly greater alignment.

For this study, we scored six state model rubrics (California, Colorado, Kentucky, Massachusetts, North Carolina, and Tennessee) and 25 district rubrics from five states (Colorado, Florida, Kentucky, Massachusetts, and Tennessee). Two of the states represented (Tennessee and North Carolina) exercise high levels of state control, two states (Colorado and Massachusetts) exercise moderate levels of state control, and three states (Florida, California, and Kentucky) exercise low levels of state control.

Table 11 shows average alignment score for all district rubrics scored, by level of state control, along with the percentage of rubrics showing various levels of alignment. The group with the highest percentage of rubrics scoring greater than .50 is the moderate state control group.

Table 11. Average Score for District Rubrics, by Level of State Control

Type	*N*	Average Alignment Score	% Rubrics Scoring > .50	% Rubrics Scoring Between .25–.49	% Rubrics Scoring < .25
Low state control	16	.16	0%	25%	75%
Moderate state control	8	.17	12.5%	0%	87.5%
High state control	1	.38	0%	100%	0%

The high state control rubric scored higher than rubrics from states with low or moderate levels of state control, on average, and there was no notable difference in average alignment score for rubrics from states with low levels of state control as compared with states with moderate levels of state control. Given the small sample size, scoring additional district-level rubrics from Tennessee or North Carolina, or from other states with high levels of state control, would help clarify if this is a valid finding.

Some evidence from this study suggests that state control may not influence alignment. The three highest scoring district rubrics, all scoring in the top 25% of rubrics scored, came from Florida (Pinellas, .38), a low control state (state model not scored); Colorado (LEAP, .75), a moderate control state (Colorado state model, .42); and Tennessee (Memphis, .38), a high control state (Tennessee state model, .44). The Memphis rubric received an alignment score similar to the Tennessee state instrument, which is to be expected in a state with a high level of state control. The LEAP rubric scored even higher than the Colorado state model, suggesting that some flexibility in a moderate control state with a strong state model has the potential to result in a district-level rubric with even greater alignment than the state model. In contrast, district-level rubrics from Massachusetts (moderate control), where the state model was less aligned, generally received lower alignment scores. Although these findings suggest that high and moderate levels of state control may contribute to higher district scores, especially in states with strong state models, further analysis is needed to test this hypothesis, given the small sample size and the strength of the Pinellas rubric from Florida, a low control state.

Alignment by Common Practice

In terms of alignment to specific common practices, rubrics from states with moderate and high levels of state control generally follow the same patterns seen across the full set of rubrics scored, with the highest alignment to Common Practices 1, 2, 5, and 8. See Table 12 for average alignment scores for each practice, by level of state control.

Table 12. Average Alignment Scores for Each Common Practice, by Level of State Control

Level of State Control	N	1. Establish Appropriate Learning Goals	2. Engage Students With Cognitively Challenging Tasks and Activities	3. Promote Strategic and Appropriate Use of Domain-Specific Tools and Resource	4. Facilitate Student Discourse	5. Engage in Purposeful Questioning	6. Integrate Key Components of the Domain	7. Promote Use of Academic and Domain-Specific Language and Vocabulary	8. Assess Important Skills and Understandings
Low	16	0.16	0.27	**0.13**	0.08	0.21	0.06	0.02	0.38
Moderate	8	0.25	0.29	0.19	0.13	0.21	**0.10**	**0.17**	0.13
High	1	**1.0**	**0.67**	0.0	**0.0**	**0. 33**	0.00	0.00	**1.00**

Rubrics from states with low levels of state control are most frequently aligned with Common Practices 2, 5, and 8 as well but show lower levels of alignment to Common Practice 1 (alignment to state standards) than seen across the full set of rubrics scored. One interpretation could be that in districts with low levels of state control, establishing appropriate learning goals receives less focus, perhaps in lieu of instructional strategies considered more attainable or tangible.

The lowest scoring rubrics come primarily from Kentucky, a low control state. Many of the district-level rubrics from this state showed no alignment to two or more practices where the state model showed at least partial alignment. In Colorado, a moderate control state, some district rubrics showed similar patterns, with no alignment to two or more practices where the state model showed at least partial alignment. A notable exception, however, is Denver's LEAP, which showed full alignment to six of the eight common practices. In contrast, the district-level rubric from Tennessee (Memphis), a high control state, showed alignment to four of the same five practices as the state model. See Appendix B for individual rubric scores, by common practice.

Variation in Alignment Within States

AIR scored at least two rubrics from five of the states included. For those states, we were able to calculate means and ranges within state and to compare individual rubric scores to the state model score (for all states but Florida). As such, we were able to begin exploring the relationship between state alignment scores, district alignment scores, and level of state control. Table 13 shows the average alignment scores, by state, along with the state model score (if available) and the range of alignment scores within each state.

Table 13. Average District Alignment Score Compared With State Model

State	Level of State Control	Number of Rubrics Scored in State	State Framework/ Model Alignment Score	Average Alignment Score	Range of Alignment Scores
TN	High	2	.44	0.41	0.38–0.44
FL	Low	3	n/a[a]	0.29	0.21–0.38
CO	Moderate	7	.42	0.22	0–0.75
KY	Low	14	.35	0.15	0–0.35
MA	Moderate	2	.19	0.15	0.06–0.19

Note. n/a = not applicable.
[a] Florida does not publish a state model rubric.

Two states stood out for their wide range of scores across district rubrics within the state: Colorado, which had a range of 0 to .75 (highest score of all rubrics), and Kentucky, which had a range of 0 to .35. Although one explanation for this may be the sample size scored in Colorado and Kentucky, preliminary analyses suggest there may be other interpretations as well. The flexibility Colorado affords its districts, as a moderate control state, appears to result in varying degrees of success of alignment. Similarly, district rubrics from Kentucky, a low control state, appear to suffer, in part because of district-level decisions to reduce the specificity of rubrics, leaving some instruments with only a scant checklist to guide teachers and their evaluators. Underscoring these findings, the Colorado state model score was .42, with LEAP (CO) scoring substantively above that at .75, whereas the Kentucky state model was scored .35, with all but one of the 13 district-level rubrics scoring below that, and six of those scoring between 0 and .10. Findings from Kentucky suggest that too little state control, especially in a state with a state model that is less aligned, can result in weaker district-level rubrics, as measured by alignment to the AIR Protocol.

Prominent Instructional Practices

Key Findings:

- More than half of teacher evaluation rubrics examined showed at least partial alignment to three of the eight common instructional practices: Common Practice 1 (Establish Appropriate Unit Learning Goals), Common Practice 2 (Engage Students With Cognitively Demanding Tasks), and Common Practice 5 (Engage in Purposeful Questioning). Forty-nine percent of rubrics showed at least partial alignment to Practice 8 (Assess Important Skills and Understandings) as well.
- Less than one quarter of rubrics scored showed at least partial alignment to three of the eight common instructional practices: Common Practice 3 (Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources), Common Practice 6 (Integrate Key Components of Content and Dispositions of the Domain), and Common Practice 7 (Promote Use of Academic and Domain-Specific Language and Vocabulary).
- On average, subject-specific rubrics (from nonprofit organizations) received higher alignment scores than non-subject-specific rubrics.

Most Commonly Observed Common Practices and General Elements

Although overall alignment to the AIR Protocol was lower than expected for general rubrics scored, the majority of rubrics did show at least partial alignment to some of the eight common practices. More than half of teacher evaluation rubrics examined showed at least partial alignment to three of the eight common instructional practices: Common Practice 1 (Establish Appropriate Unit Learning Goals), Common Practice 2 (Engage Students With Cognitively Demanding Tasks), and Common Practice 5 (Engage in Purposeful Questioning). Forty-nine percent of rubrics showed at least partial alignment to Common Practice 8 (Assess Important Skills and Understandings) as well. Each practice is made up of one or more specific elements, and within the practices most commonly observed, rubrics were more frequently aligned to some specific elements of Common Practices 1, 2, 5, and 8 than others. These trends are described in more detail here.

Common Practice 1. The most commonly observed element across the 37 general rubrics scored was *Align to the state standards for students at that grade/level* (Common Practice 1), which appeared in 19 rubrics. Given that aligning learning goals with standards has been common practice since the standards movements of the 1990s, it may not be surprising that most rubrics aligned with this element. The other element in Common Practice 1, however, related to establishing a coherent progression of content and skills, was not as frequently observed (seven rubrics). The lack of alignment to this element suggests that rubrics may not be sufficiently focusing on coherence, as called for by the Common Core State Standards.

Common Practice 2. Two of the most frequently observed elements came from Common Practice 2: *Think deeply about the material (e.g., synthesize information/concepts), reason about the content, or make connections across topics* (16 rubrics) and *Engage with or apply their understanding and skills to challenging or complex tasks or texts* (14 rubrics). Challenging students has long been a recognized and desired goal for educators, so it is encouraging to see relatively high levels of alignment to these elements across the rubrics scored. Additional rubrics may have been considered aligned to these elements except for the lack of explicit reference to the source of challenge, such as specific tasks on which students work or texts they read. Rubrics showed alignment to the third element, *Engage in critical and evaluative analysis*, much less frequently, indicating that this level of expectation for student thinking is not yet widely recognized as an important feature of instruction and learning or that this level of specificity is lacking from many rubrics.

Common Practice 5. Sixteen rubrics showed alignment to the element, *Advance student reasoning and understanding (i.e., through higher order questioning)*. Many other rubrics scored mentioned higher order questions, but oftentimes the focus on higher order questions was the maintenance of rigor or high standards for student achievement. To be considered aligned to the AIR Protocol, there needed to be an indication that questioning was to be used to advance student thinking rather than just measure student thinking.

Common Practice 8. Almost all rubrics mentioned assessment in some way and nearly half (18) were aligned with the assessment practice element of Common Practice 8, *Determine student progress in developing the important skills and understandings characteristic to the domain*. To be considered aligned to this element, the rubric must specify important content, which might be the standards or "foundations" and a reference to student progress, or something similar.

Table 14 shows the five most frequently aligned general elements, from general rubrics scored, along with an example of what alignment looked like for each common element. Of particular note is that the practice of students using evidence to justify their answers—such as a textual interpretation in ELA or a problem solution in mathematics—is generally underemphasized in both responses to the teacher's questions and in discussions with peers. This lack of emphasis on evidence in most rubrics is notable given the emphasis it receives in the standards themselves.

Table 14. Most Frequently Found Instructional Elements and Rubric Examples

Instructional Practice	General Element	Count (37 total)
1. Establish Appropriate Unit Learning Goals	Align to the state standards for students at that grade or level.	19
Example: Colorado State Model	ELEMENT A: Teachers provide instruction that is aligned with the Colorado Academic Standards; their district's organized plan of instruction; and the individual needs of their students.	
8. Assess Important Skills and Understandings	Determine student progress in developing the important skills and understandings characteristic to the domain.	18
Example: Tennessee TEAM	Assessment plans: ■ Are aligned with state content standards; ■ Have measurement criteria; ■ Measure student performance in more than two ways (e.g., in the form of a project, experiment, presentation, essay, short answer, or multiple choice test); ■ Require written tasks; and ■ Include performance checks throughout the school year.	
5. Engage in Purposeful Questioning	Advance student reasoning and understanding (i.e., through higher order questioning).	16
Example: Pike County, KY	Questions require extended thinking and promote application of concepts or skills (i.e., common verbs—design, connect, critique, analyze, create, prove).	
2. Engage Students With Cognitively Challenging Tasks and Activities	Think deeply about the material (e.g., synthesize information or concepts), reason about the content, or make connections across topics.	15
Example: NC State Model	4e. Teachers help students develop critical-thinking and problem-solving skills. Teachers encourage students to ask questions, think creatively, develop and test innovative ideas, synthesize knowledge, and draw conclusions. They help students exercise and communicate sound reasoning; understand connections; make complex choices; and frame, analyze, and solve problems.	
2. Engage Students With Cognitively Challenging Tasks and Activities	Engage with or apply their understanding and skills to challenging or complex tasks or texts.	14
Example: Silverton, CO	Identify compelling topics that ask students to learn about and analyze sophisticated situations.	

Least Commonly Observed Common Practices and General Elements

Less than 25% of general rubrics scored showed at least partial alignment to three of the eight common instructional practices: Common Practice 3 (Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources), Common Practice 6 (Integrate Key Components of Content and Dispositions of the Domain), and Common Practice 7 (Promote Use of Academic and Domain-Specific Language and Vocabulary). Within the practices least commonly observed, rubrics were least frequently aligned to some specific elements of Common Practices 3, 6, and 7.

Common Practice 3. Although 24% of rubrics referred to student use of tools and resources, only one of the rubrics scored explicitly referred to students making choices about which tool to use. Although a number of rubrics described teacher use of technology, few rubrics mentioned student use of technology.

Common Practice 6. Common Practice 6 is defined for general rubrics in the AIR Protocol by one element, Attends to and integrates key components, proficiencies, and dispositions within the discipline. Although content was addressed in all rubrics and integration or equivalent constructs for content understanding were addressed in many rubrics, only two rubrics specifically included instruction focusing attention on practices or dispositions within the discipline.

Common Practice 7. All three elements comprising Common Practice 7, Promote Use of Academic and Domain-Specific Language and Vocabulary, were among the five least commonly observed elements. Teacher use of correct language and vocabulary appeared in some rubrics and others included general academic language, but explicit statements about encouraging students to use domain-specific language, active attention to domain-specific language, and relating formal language to students' informal language were rare. Although the phrase "academic language" appeared frequently across rubrics, attention to vocabulary was not often present.

The focus on content to the exclusion of ways of working, habits of mind, or dispositions (Common Practice 6), and using the language (Common Practice 7) and tools (Common Practice 3) of a discipline reflects the traditional view of standards as describing a body of things to be known and skills that one will be able to do. If teachers are to be encouraged to instruct students in the use of discipline-specific tools, connect discipline content standards with practices of a discipline, and help students develop the vocabulary and language structures of a discipline, rubrics must explicitly include these expectations.

Perhaps not surprisingly, alignment to all elements of a practice within a rubric was rarely observed. For instructional practices with multiple elements, Common Practice 1 (Establish Appropriate Unit Learning Goals) yielded the highest rate of full alignment, with 19% of rubrics scored showing alignment to all elements in the instructional practice.

Table 15 shows the least frequently aligned general elements, from general rubrics scored, along with an example of what alignment looked like in the rubrics where alignment to these elements was observed.

Table 15. Least Frequently Found Instructional Elements and Rubric Examples

Instructional Practice	General Element	Count (37 total)
7. Promote Use of Academic and Domain-Specific Language and Vocabulary	Asks students to use academic and domain-specific vocabulary.	3
Example: Colorado LEAP	INDICATOR I.4: Ensures all students active and appropriate use of academic language. ■ Consistently and explicitly teaches and models precise academic language connected to the content-language objective(s) using the target language (students' Language 1 or 2, as appropriate). ■ Provides frequent opportunities within the content for students to use academic language in rigorous, authentic ways through listening, speaking, reading, and writing. ■ Acknowledges students' use and attempts at using academic language to develop concepts, and coaches students when academic language is not used or is used incorrectly. ■ Consistently encourages students to use complete sentences. ■ Students use academic language (in their native language or English) with the teacher, peers, and in their writing. ■ Students are observed using target language in a variety of contexts and for cognitively demanding tasks, often in collaboration with other students. ■ Students regularly and accurately use content vocabulary and language forms relevant to the objective(s).	
7. Promote Use of Academic and Domain-Specific Language and Vocabulary	Demonstrates and draws attention to correct use of academic and domain-specific vocabulary.	3
Example: Brevard, FL	VI. Models and teaches clear, acceptable communication skills. 1. Directions, procedures, and feedback are clear to students. 2. Teacher's spoken and written language conform to standard English. 3. Teacher uses academic language and content vocabulary accurately.	

Instructional Practice	General Element	Count (37 total)
Example: Colorado LEAP	INDICATOR I.4: Ensures all students active and appropriate use of academic language. ■ Consistently and explicitly teaches and models precise academic language connected to the content-language objective(s) using the target language (students' Language 1 or 2, as appropriate). ■ Provides frequent opportunities within the content for students to use academic language in rigorous, authentic ways through listening, speaking, reading, and writing. ■ Acknowledges students' use and attempts at using academic language to develop concepts, and coaches students when academic language is not used or is used incorrectly. ■ Consistently encourages students to use complete sentences. ■ Students use academic language (in their native language or English) with the teacher, peers, and in their writing. ■ Students are observed using target language in a variety of contexts and for cognitively demanding tasks, often in collaboration with other students. ■ Students regularly and accurately use content vocabulary and language forms relevant to the objective(s).	
6. Integrate Key Components of the Domain	Attends to and integrates key components, proficiencies, and dispositions within the discipline.	2
Example: McLean County Kentucky	3.2: Develops instruction that requires students to apply knowledge, skills and thinking processes. 3.3: Integrates skills, thinking processes, and content across disciplines.	
3. Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources	Student involvement in a decision-making process through which to determine which tool is appropriate to the work.	1
Example: Colorado LEAP	INDICATOR I.2: Provides rigorous tasks that require critical thinking with appropriate digital and other supports to ensure students' success. Provides digital resources or tools as a support for rigorous tasks when appropriate. Students (including students of color, linguistically diverse students, those with disabilities and those identified as gifted and talented) execute increasingly complex tasks by formulating hypotheses, analyzing data, or solving real-world problems to deepen their understanding of the content-language objective(s).	
7. Promote Use of Academic and Domain-Specific Language and Vocabulary	Connects students' informal language use with formal language use.	1

Instructional Practice	General Element	Count (37 total)
Example: Colorado LEAP	INDICATOR I.4: Ensures all students are active and appropriate use of academic language. ■ Acknowledges students' use and attempts at using academic language to develop concepts, and coaches students when academic language is not used.	

Subject-Specific Rubrics

The subject-specific rubrics scored had a higher level of alignment with the subject-specific elements of the common instructional practices than the general rubrics had to the general elements of the common instructional practices. The lowest score of alignment seen in a subject-specific rubric was .22—higher than half of the general rubrics—and the mean score of the subject-specific rubrics was .38, compared with a mean of .21 for the general rubrics. However, three of the 37 general rubrics (Colorado LEAP, Alliance College-Ready, and TNTP) had a higher overall average alignment score than any of the eight subject-specific rubrics.

Alignment to Common Practices and General Elements

The three most frequently aligned common instructional practices for general rubrics, Common Practice 1, Establish Appropriate Unit Learning Goals (averages score .51 across all rubrics), Common Practice 2, Engage Students With Cognitively Demanding Tasks (averages score .51 across all rubrics), and Common Practice 5, Engage in Purposeful Questioning (averages score .51 across all rubrics) also were the most frequently aligned common instructional practices in the subject-specific rubrics. The relative frequency was higher in the subject-specific rubrics, with each of the three practices seen in 88% of rubrics.

Alignment to at least some specific elements of Common Practice 4 (Facilitate Student Discourse) was much higher for the subject-specific rubrics (average score .88 across all rubrics) than for the general rubrics (average score .30).

Alignment of subject-specific rubrics to Common Practices 3, 6, 7, and 8 varied by subject:

- For Common Practice 3 (Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources), mathematics-specific rubrics (average .50) were aligned more frequently than the general rubrics (24%), whereas the ELA-L–specific rubrics showed no alignment.
- For Common Practices 6 and 7, Integrate Key Components of Content and Dispositions of the Domain and Promote Use of Academic and Domain-Specific Language and Vocabulary, 100% of mathematics rubrics were aligned to at least one of the mathematics-specific elements as compared with only 25% of ELA-L rubrics to at least one ELA-L–specific element. Only 5% and 14% of general rubrics were aligned to the general elements for Practices 6 and 7, respectively.
- Elements of Common Practice 8 (Assess Important Skills and Understandings), seen in 49% of the general rubrics, were seen in 100% of the ELA-L–specific rubrics but none of the mathematics specific rubrics.

The greater frequency of alignment seen with mathematics-specific rubrics for practices related to use of discipline-specific tools (Common Practice 3), integration of content and dispositions (Common Practice 6), and domain-specific language and vocabulary (Common Practice 7) may be a reflection of the explicit inclusion of the mathematical practices at the beginning of the mathematics standards. Mathematical Practice 6 is about precise language, Mathematical Practice 7 concerns the strategic use of tools, and there is an explicit statement about connecting "the mathematical practices to mathematical content in mathematics instruction," which seems related to the higher rates of alignment with elements of Instructional Practice 6 of the AIR Protocol.

Alignment to Subject-Specific Practices and Elements

Subject-specific elements were written for ELA-L and for mathematics for both the common (across subjects) instructional practices and for additional instructional practices pertaining only to those subjects. The number of additional practices and elements for each subject area is summarized in Table 16 and displayed in Tables 2 and 3.

Table 16. Subject-Specific Practices and Elements

Subject	Additional Subject-Specific Practices	Additional Elements (Subject-Specific Practices)	Additional Subject-Specific Elements (Common Practices)
ELA-L	3	9	23
Mathematics	4	9	23

Subject-Specific Practices. Of the three ELA-L–specific instructional practices, two of four rubrics were aligned to the one element that makes up ELA-L 1, Early Foundational Reading practice; one of four was aligned to all elements of ELA-L 2, Writing for Research, Argumentation, and Narrative practice, none of the other three rubrics was aligned to any element of the practice; and one rubric was aligned to all four elements of ELA-L 3, Building Independence for All Students Practice, while the other three were each aligned to two of the four elements for the practice.

For the mathematics-specific instructional practices, one of four rubrics showed alignment to one element of Math 4, Encourage Abstract Reasoning practice, and one of four rubrics showed alignment to elements from Math 1, Real-World Contexts and Modeling practice. Two of four rubrics emphasized use of Math 3, mathematical representation (one element of this practice), but none of them was aligned to teachers making explicit the way representations relate to each other. Similarly, all four rubrics were aligned to opportunity for challenge (one element of this practice), but none showed alignment to teacher support for students as they struggle with mathematics.

Subject-Specific Elements. All of the nine elements written for the three ELA-L–only instructional practices were seen in at least one rubric, with 36% elements represented, on average, in each rubric. This data seem to indicate some common understandings across the field. Of the 32 total subject-specific elements for ELA-L, including the 23 comprising common practices and the nine comprising ELA-L–only practices, two elements were seen in all four ELA-L–specific rubrics and six were seen in three of the four rubrics. Nine elements were absent from all four evaluation rubrics examined.

None of the ELA-L–specific rubrics were aligned to the ELA-L–specific elements for Common Practice 3 (Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources). Only one of four ELA-L–specific rubrics was aligned to element of Common Practice 7 (Promote Use of Academic and Domain-Specific Language and Vocabulary), and that rubric was only aligned to one of four elements for the practice. One rubric showed alignment with the single element for Common Practice 6 (Integrate Key Components of the Domain).

Overall, the average alignment of the rubrics to the elements of the ELA-L–specific instructional practices (44%) is higher than the alignment to the ELA-L–specific elements for the general instructional practices (33%).

For mathematics-specific rubrics, 38% of elements were represented, on average, in each rubric. Four of the nine elements written for the mathematics-only instructional practices were not seen in any of the rubrics, three elements were seen in one rubric, one was seen in two rubrics, and only one, *Provides opportunities for students to engage with challenging mathematics material*, was seen in all four rubrics. The general nature of this element and the absence of its companion element, *Supports students as they grapple with challenging material without immediately "doing it for them,"* coupled with the low rates of alignment of the elements, indicate that these specific "look-fors" have not been widely identified as important. Of the 32 total subject-specific elements for mathematics, including the 23 comprising common practices and the nine comprising mathematics-only practices, three elements were seen in all four mathematics-specific rubrics, and seven were seen in three of the four rubrics. Eleven elements were not considered aligned with any of the four evaluation rubrics examined.

None of the mathematics-specific rubrics was aligned to any of the mathematics-specific elements for Common Practice 8 (Assess Important Skills and Understanding). At least two rubrics were aligned to one or more elements for each of the other practices.

Overall, the average alignment of the rubrics to the elements of the mathematics-specific instructional practices (25%) is lower than the alignment to the mathematics-specific elements for the general instructional practices (43%). This is the opposite of the finding for ELA-L.

Table 17 summarizes the number and percentage of rubrics in each category that contained at least some aspects of each instructional practice. Although the subject-specific sample is smaller, several practices, especially those representing cognitive demand and specific instructional moves (including cognitively demanding tasks, student discourse, and purposeful teacher questioning) were more likely to be represented in the subject-specific group. Integrating practices of the academic domain (e.g., teaching fluency and conceptual understanding) and use of academic language were stressed most often in mathematics instruments.

Table 17. Number and Percentage of Rubrics Aligned to Each Practice, by Category

Instructional Practice	Number of Rubrics Aligned With at Least One Element of the Instructional Practice (Percentage)		
	ELA-L–Specific (*n* = 4)	Math-Specific (*n* = 4)	General (*n* = 37)
1. Establish Appropriate Unit Learning Goals	3 (75%)	3 (75%)	19 (51%)
2. Engage Students With Cognitively Demanding Tasks	4 (100%)	3 (75%)	19 (51%)
3. Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources	0 (0%)	2 (50%)	9 (24%)
4. Facilitate Student Discourse	3 (75%)	4 (100%)	11 (30%)
5. Engage in Purposeful Questioning	4 (100%)	3 (75%)	19 (51%)
6. Integrate Key Components of Content and Dispositions of the Domain	1 (25%)	4 (100%)	2 (5%)
7. Promote Use of Academic and Domain-Specific Language and Vocabulary	1 (25%)	4 (100%)	7 (14%)
8. Assess Important Skills and Understandings	4 (100%)	0 (0%)	18 (49%)

Conclusion

Teaching is a complex profession. Teacher evaluation rubrics should be tools to assess how well teachers are going about the complex tasks associated with instructing students and offer them feedback for improving their practice. Most instruments, however, currently fail to offer teachers substantive guidance. Rubrics focus on noninstructional dispositions, contain few specifics, and often ignore practices connected to the particular subjects and standards that teachers are asked to teach. Too many instruments seem devoted to creating a universal description of good teaching at the expense of providing real guidance for the many kinds of instruction that take place in a typical Grades K–12 system. States and districts cannot hope to substantively change instruction with generic, uniform rubrics that contain significant amounts of noninstructional content. As others have observed, "In the desire to create one-size-fits-all systems of teacher evaluation, states and districts may risk relegating the substance of instruction to the sidelines" (Hill & Grossman, 2013, p. 380).

If there is an overarching takeaway from this research, it is that some tailoring of teacher evaluation rubrics is necessary to achieve alignment to goals of subject-specific rigor embedded in new college- and career-ready standards. In particular, based on what we learned from studying 45 rubrics currently being used to evaluate teachers and guide their instruction, the field should consider additional instructional content in teacher evaluation rubrics, including additional subject-specific content that will provide more specific guidance and direction to teachers as they attempt to implement instruction

that will help their students achieve the standards. This research does not unpack how to balance this specificity with local flexibility and autonomy for teachers to innovate and take ownership of their practice. It does, however, highlight some of the more successful rubrics in the field and provide some steps that can be taken to improve existing instruments. The current lack of clear definition or sufficient detail means that teachers may neither be rewarded for aligned practices nor sufficiently guided in how to improve their teaching.

We recommend three key ways to revise teacher evaluation rubrics to more closely align with the new student learning standards.

First, most of the rubrics assessed here should include a broader array of instructional practices in order to align with the expectations of the Common Core. A number of instruments contain a notable amount of noninstructional content at the expense of instructionally focused content. Key instructional moves such as emphasizing academic vocabulary, employing technology in the classroom, engaging students in rigorous peer-to-peer discourse, or requiring students to substantiate answers appear infrequently. In other cases, practices included in the AIR Protocol were alluded to but not clearly stated or described in a way that might offer feedback to teachers. The range of sensitivity scores presented in Appendix C illustrates the challenge coders faced in agreeing on the presence of certain, specific practices that were not well defined in the teacher rubrics. For example, the nature of cognitively demanding tasks, where students engage in some kind of critical analysis of a text or solution to a problem, proved to be only vaguely defined for most teachers and evaluators.

Second, teacher evaluation rubrics should either be more subject-specific than most currently are or contain broad, generalizable instructional principles with more subject-specific subindicators. The generalized, uniform rubrics that dominate the field currently offer less substantive instructional guidance than do those that contain content specific to subjects, grade levels, or both. The results of this analysis, as well as those from AIR's previous reports to the Gates Foundation in May and December 2015 (see Welch, Freed et al., 2015; Welch, Potemski et al., 2015), indicate that teacher evaluation rubrics with at least some subject-specific content—if not distinct instruments for different subject areas—score more favorably in terms of their alignment to Common Core–aligned teaching practices.

The authors of the AIR Alignment Protocol see the Common Core State Standards as two sets of standards rather than one, emphasizing certain knowledge and skills that are particular to their subject area domains. Each of the common practices can be defined in terms of particular elements, or specific instructional moves, that are particular to ELA-L and mathematics. In light of this perspective, it is unsurprising that rubrics like those from SAP and the ELA-L-specific version of Danielson's FFT scored well using the subject-specific standard. But general rubrics can achieve similar clarity as well. The state model rubric from Colorado is an example of an instrument that might be created to highlight certain universally desirable teaching practices for all teachers but still use subindicators specific to particular subject areas or grade levels. Teachers, after all, often specialize in particular subjects, teaching only one or a selected few disciplines. Even in self-contained classrooms, teachers often devote separate blocks of time to subject-specific lessons and activities. The content standards that teachers are asked to teach are published separately by subject. Students are assessed and graded

separately by subject. Ideally, teacher evaluation rubrics should align with these particular expectations and incentivize subject-specific teaching practices.

Third, at the state level, these results suggest that policymakers would achieve greater alignment with sound state models that offer ample, clear detail on desired practices while considering ways to balance guidance with local flexibility. Other research in implementing change has found that engagement and flexibility for local actors are important and complementary components for new educational initiatives (Fullan, 2007). Districts and schools can be engaged in the process of creating the guiding framework that aligns curriculum, instruction, and evaluation, adding specificity to broadly defined tools like rubrics and helping find ways to make curricular demands and evaluation principles work together. Results from this study suggest that some, but not total, flexibility in implementation may be most effective with some guidelines and an effective state model. Again, instruments from Colorado offer some examples of this kind of moderated state control in practice, with the state model offering some subject-specific practices and Denver's LEAP rubric touching on the majority of the instructional practices in the AIR Protocol. States may not have the capacity or desire to develop a large number of different instruments for teachers of different subject or grades. Further, states and districts also may find it desirable to publish universal guidelines for teachers to promote uniform expectations for the profession while leaving the assessment of particular instructional moves within these principles to district evaluators, school leaders, or department heads. Some systems, for example, have offered districts localized feedback on whether instruments meet a state's best practices (see Jacques, 2013). In short, guidance and support may yield greater alignment than just mandating a rubric or allowing districts to create their own.

When considering the results of this analysis, it is important to consider the wide array of instruments in use in the field of teacher evaluation and the many kinds of instructional practices, background knowledge, and professional dispositions emphasized by them. Further, because most instruments were written to apply to all teachers in all grades, most contain broad, general wording that not only challenged the reliability of the coding teams but also would likely make it difficult to provide specific, actionable feedback to teachers. In general, policymakers would be advised to focus teacher evaluation rubrics more on classroom instruction. More specifically, teacher rubrics lack the required focus on the particular instructional moves relevant to the teacher being assessed (i.e., appropriate for grade level and subject being taught) in order to provide more specific guidance to teachers that is coherent with the demands of state student learning standards.

Most teachers in states with Common Core–aligned practices will hear that their classroom instruction should conform to these norms. Teacher evaluation rubrics should be strengthened by emphasizing particular instructional practices that align with these standards and the curricular materials that teachers will encounter going forward. Clear, specific feedback, grounded in what and how teachers teach, is the only way teacher evaluation materials will lead to improved instruction and not just another handed-down policy that makes teaching additionally complex.

References

Cantrell, S., & Kane, T. J. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study* (Policy and Practice Brief). Seattle, WA: Bill & Melinda Gates Foundation.

Center on Great Teachers and Leaders. (2013). *Databases on state teacher and principal evaluation.* Retrieved from http://resource.tqsource.org/stateevaldb/StateRoles.aspx

Cohen, D. K. (2010/2011). Learning to teach nothing in particular: A uniquely American educational dilemma. *American Educator, 34*(4), 44–46.

Curtis, R. (2012). *Building it together: The design and implementation of Hillsborough County Public Schools' teacher evaluation system.* Washington, DC: The Aspen Institute.

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement.* New York, NY: Teachers College Press.

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research, 53*(3), 285-328.

Donaldson, M. L., Cobb, C., LeChasseur, K., Gabriel, R., Gonzales, R., Woulfin, S., & Makuch, A. (2014). *An evaluation of the pilot implementation of Connecticut's system for educator evaluation and development.* Storrs, CT: Center for Education Policy Analysis.

Elmore, R. E. (2005). *School reform from the inside out: Policy, practice, and performance.* Cambridge, MA: Harvard Education Press.

Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher, 43,* 100–107.

Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York, NY: Teachers College Press.

Hayes, L., & Lillenstein, J. (2015). *A framework for coherence: College and career readiness standards, multi-tiered systems of support, and educator effectiveness.* Washington, DC: Center on Great Teachers and Leaders.

Hill, H. C., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 371–384.

Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher, 33*(8), 16–30.

Insight Core Education Group. (2013). *Insight core framework rubric.* Retrieved from http://www.insighteducationgroup.com/insight-core-framework-landing-page

Interstate Teacher Assessment and Support Consortium. (InTASC). (2013). *InTASC model core teaching standards and learning progressions for teachers 1.0.* Washington, DC: Council of Chief State School Officers. Retrieved from http://www.ccsso.org/Documents/2013/2013_INTASC_Learning_Progressions_for_Teachers.pdf

Jacques, C. (2013). *Leveraging teacher talent: Peer observation in educator evaluation.* Washington, DC: Center on Great Teachers and Leaders.

Leo, S. F., & Coggshall, J. G. (2013). *Creating coherence: Common core state standards, teacher evaluation, and professional learning.* Washington, DC: Center on Great Teachers and Leaders.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education, 79*(4), 33–53.

National Council of Teachers of Mathematics. (2013). *Principals to action.* Reston, VA: Author. Retrieved from http://www.nctm.org/uploadedFiles/Standards_and_Positions/PtAExecutiveSummary.pdf

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis, 23*(4), 297–321.

Polikoff, M. S., & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis, 36*(4), 399–416.

Porter, A. (1995). The uses and misuses of opportunity-to-learn standards. *Educational Researcher, 24*(1), 21–27.

Smith, M. S., & O'Day, J. (1991). *Putting the pieces together: Systemic school reform* (CPRE Consortium for Policy Research in Education Policy Brief No. RB-06). Philadelphia: University of Pennsylvania Graduate School of Education.

Smylie, M. A. (2014). Teacher evaluation and the problem of professional development. *Mid-Western Educational Researcher, 26*(2), 97–111.

Student Achievement Partners. (2013a). *Instructional practice guide: ELA, K–2.* New York, NY: Author.

Student Achievement Partners. (2013b). *Instructional practice guide: ELA, 3–5.* New York, NY: Author.

Student Achievement Partners. (2013c). *Instructional practice guide: ELA, 6–12.* New York, NY: Author.

Student Achievement Partners. (2013d). *Instructional practice guide: Mathematics, high school.* New York, NY: Author. Retrieved from http://achievethecore.org/content/upload/InstructionalPracticeGuide_MATH_HS_D_09192013.pdf

Student Achievement Partners. (2013e). *Instructional practice guide: Mathematics, K–8.* New York, NY: Author.

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628–3651.

Teaching Works. (2012). *High leverage practices.* Ann Arbor, MI: University of Michigan. Retrieved from http://www.teachingworks.org/work-of-teaching/high-leverage-practices

TNTP. (2014). *TNTP core teaching rubric.* Brooklyn, NY: Author. Retrieved from http://tntp.org/publications/view/tntp-core-teaching-rubric-a-tool-for-conducting-classroom-observations#download

Tri-State Collaborative. (2013a). *Tri-State ELA rubric.* Washington, DC: Achieve. Retrieved from https://www.engageny.org/file/536/download/tri-state-ela-rubric.pdf?token=mLnNk7dj6A6MoEmy7OUn_Ut9CWPozcyMHwGHdVxnwuQ

Tri-State Collaborative. (2013b). *Tri-State ELA rubric K–2.* Washington, DC: Achieve. Retrieved from https://www.engageny.org/file/7706/download/tri-state-ela-rubric-k-2.pdf?token=UYnCxoZ_qcsXcXhes1ccsi_MoY5FMU3r3_u02Kwui2k

Tri-State Collaborative. (2013c). *Tri-State math rubric.* Washington, DC: Achieve. Retrieved from https://www.engageny.org/file/2761/download/tri-state-math-rubric_0.pdf?token=VSKaQ7J17LJjhrYU1yVRTVA1ObUituh9qe4BKWYx3Nc

Welch, M., Freed, D., Jenuwine, H., Johnston, W. T., Smith, T., & Isaia, R., Booth, S. R. (2015). *Evaluation rubrics' alignment to the core: Assessing alignment in district and other teacher evaluation rubrics.* Washington, DC: American Institutes for Research.

Welch, M., Potemski, A., Freed, A., Jacques, C., Smith, T., & Isaia, R. (2015). *Teacher evaluation and teaching to the Core: Measuring alignment between policies on teachers and teaching.* Washington, DC: American Institutes for Research.

Whitehurst, G. J. R., Chingos, M. M., & Lindquist, K. M. (2014). *Evaluating teachers with classroom observations: Lessons learned in four districts.* Washington, DC: Brown Center on Education Policy at Brookings.

Appendix A:
Teacher Evaluation Instrument Alignment Protocol

Protocol of Instructional Practices That Support Common Core–Aligned State Standards Alignment Ratings for [STATE]

State's Overall Rating of Alignment of Teacher Evaluation Criteria to Practices That Support Common Core–Aligned State Standards

State Subtotals

[NAME of State]: **State's Score on Common Instructional Practices**							
Establish Appropriate Learning Goals	Engage Students With Cognitively Challenging Tasks and Activities	Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources	Facilitate Student Discourse	Engage in Purposeful Questioning	Integrate Key Components of the Domain	Promote Use of Academic and Domain-Specific Language and Vocabulary	Assess Important Skills and Understandings

Common Mathematics Alignment	Common ELA-L Alignment

Mathematics-Specific Practices				ELA-L–Specific Practices		
Embed Mathematics in Real-World Contexts	Support Students' Productive Struggle	Promote Use of Multiple Representations	Encourage Abstract Reasoning	Early Reading Foundational Skills	Writing for Research, Argumentation, and Narrative	Building Independence for All Students

Instructions in the Use of the AIR Protocol

The American Institutes for Research (AIR) Alignment Protocol (AIR Protocol) is intended to measure the degree of alignment between teacher evaluation rubrics and instructional practices that support Common Core–aligned standards. In applying the AIR Protocol to an evaluation instrument, coders should consider whether each indicator on the selected teacher evaluation rubric aligns to each element of instructional practice presented here. The scoring formulas award an evaluation rubric a full point for each subject-specific element and a half point for each generic element.

Two coders use the AIR Protocol to score each teacher evaluation instrument: one specializing in English, language arts, and literacy, and one in mathematics. The accompanying scoring matrix will calculate and total scores for each Common Practice and each of the subject-specific practices particular to each coder's area of expertise.

Rules on Use of the AIR Protocol

1. Read the original teacher evaluation instrument, considering which standards apply largely or even partially to each instructional practice and its constituent elements.
    a. Coders should consider each individual standard and subindicator on the rubric in question for alignment with the elements of the AIR Protocol. In considering alignment, coders may read the standard, its component indicators, any text that describes or introduces the standard, subelements or indicators, and descriptors or attributes. *Do not* consider any listed examples.
2. Read each indicator within each standard on the teacher evaluation rubric. Coders should read the general description of the indicator and then **the minimally acceptable or passing level for meeting that standard in the evaluation rubric** (e.g., *proficient* in Colorado, or *accomplished* in West Virginia). Note that some states, such as North Carolina, use a cumulative rating system (i.e., one where each level presumes the components of the previous levels); if no level is specified, alignment to the general descriptor is acceptable. Read across the scoring matrix and determine whether each indicator is aligned with each element of the AIR Protocol. Aligned elements receive a 1 in the cross-referenced cell; nonaligned elements receive a 0.
3. The AIR Protocol is designed to focus on instructional practice. Components of teacher evaluation rubrics that focus on noninstructional qualities of teacher practice may be read for possible alignment (i.e., teacher knowledge), but alignment will often be weak.
    a. Common examples are teachers' knowledge and dispositions. Indicators within these standards can be considered aligned to AIR Protocol elements if individual indicators reference instructional practices.
    b. Coders are looking for some explicit connection to particular elements within an instructional practice. Language can be slightly different, but there needs to be some definition for a state's indicator and not just a vague title that represents a similar approach to *instruction*.
4. An indicator from a teacher evaluation instrument can be coded on more than one element of the AIR Protocol. For example, a given indicator may evaluate a teacher's design of cognitively demanding activities as well as teachers' engagement of student discourse.
5. A coder may consider multiple indicators from a state evaluation rubric when deciding if a given element from the AIR Protocol is present or not. For example, a given AIR element may have several components that are only found across multiple indicators in the rubric being coded. Alternately, a rubric may have a long collection of descriptors not tied to any one indicator. The coder should then note alignment to the selected AIR element in all applicable cells in that column.
6. In case of subject-specific elements—or subject-specific practices—a universal (or non-subject-specific) evaluation instrument must explicitly refer to that subject area (i.e., either ELA-L or mathematics). The evaluation instrument need not have an entire section or separate rubric devoted to that subject, but it **must make specific reference in the standard or indicator to the subject area in question, and not just the skills emphasized in the subject-specific elements**. The teacher evaluation rubric can make reference to subject-specific skills such as writing with evidence to earn a point for a subject-specific element. Literacy across other content areas—when specifically addressed—may qualify.
    a. For example, if a teacher evaluation instrument called for "Teachers [to] help students develop critical-thinking and problem-solving skills," such a rubric might be eligible to earn points for alignment to the general elements of Common Practice 2 (Engage Students With Cognitively Challenging Tasks and Activities) but not the mathematics-specific elements.

b. Subject-specific examples not listed in the definition of the indicator should not be counted toward subject-specific coding.

c. If a rubric is devoted entirely to a single subject area, all standards and elements of that rubrics should first be considered for subject-specific alignment but may be considered for general alignment, based on how they are worded.

d. If an indicator is coded as being subject-specific, it should not also be simultaneously coded as being aligned to a general element as well. An individual indicator may, however, be coded as aligned to multiple elements on the AIR Protocol.

7. In the case of standards alignment (Common Practice 1), it is sufficient for an evaluation instrument to specify alignment with state standards, even if the state in question is not Common Core–aligned. The evaluation instrument's indicator should explicitly call for alignment.

8. When rating for coherence, discussing sequence or prior knowledge alone is insufficient. The indicator must make reference to logic or appropriate sequence for concepts and tasks.

Note. Any reference to state standards in the document pertains to a state's standards in ELA-L and mathematics, which are presumed to be aligned with the Common Core in ELA-L and mathematics

A. Common Practices That Apply to Both Mathematics and English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects

1. Establish Appropriate Unit Learning Goals

This practice focuses on the establishment of learning goals that are aligned to the state standards and represent a purposeful leaning progression, or sequencing of concepts and material.

Establish Appropriate Unit Learning Goals		
GENERAL:		
Teacher's instruction has clearly established unit learning goals that do the following:	Present	Not Present
Align to the state standards for students at that grade or level.	1	0
Represent a coherent progression of content, skills, and materials or resources over time.	1	0
MATHEMATICS:		
Teacher's instruction has clearly established unit learning goals that do the following:	Present	Not Present
Align to state mathematics standards for students at that grade or level.	1	0
Address mathematical processes and practices (e.g., problem solving, reasoning).	1	0
Build on previously addressed concepts in logical development of the new mathematical concepts and procedures.	1	0

Establish Appropriate Unit Learning Goals		
ELA-L: **Teacher's instruction has clearly established unit learning goals that do the following:**	Present	Not Present
Align to state ELA-L standards for students at that grade or level.	1	0
Identify ELA-L skills, concepts, and content for each unit and lesson within the unit.	1	0
Align with preceding and succeeding skills and concepts within the appropriate learning progression for each unit and lesson within the unit.	1	0

2. Engage Students With Cognitively Demanding Tasks

This practice stresses instructional approaches that provide students with opportunities to engage with tasks that are challenging and require reasoning and higher level thinking.

Engage Students With Cognitively Challenging Tasks and Activities		
GENERAL: **Teacher's instruction requires students to do the following:**	Present	Not Present
Think deeply about the material (e.g., synthesize information and concepts), reason about the content, or make connections across topics.	1	0
Engage with or apply their understanding and skills to challenging or complex tasks or texts.	1	0
Engage in critical and evaluative analysis.	1	0
MATHEMATICS: **Teacher's instruction engages students with cognitively challenging tasks and activities that require students to do the following:**	Present	Not Present
Critique mathematical explanations, arguments, or solutions.	1	0
Reason and make connections within mathematics.	1	0
Compare and contrast mathematical solutions, methods, and concepts.	1	0
Engage with or apply learning to complex mathematics problems or tasks.	1	0
ELA-L: **Teacher's instruction engages students with cognitively challenging tasks and activities that require students to do the following:**	Present	Not Present
Analyze text content, craft, and structure.	1	0
Build content knowledge from a balance of informational and literary texts and tasks.	1	0
Analyze claims and evidence in argumentation.	1	0
Comprehend and critique increasingly complex literary and informational texts at or above grade level.	1	0

3. Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources

This practice emphasizes the use of tools that are germane to the discipline and student selection of those tools.

Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources		
GENERAL: **Teacher's instruction promotes the following:**	Present	Not Present
Strategic use of tools or resources, including technology, to support work in the domain.	1	0
Student involvement in a decision-making process to determine which tool is appropriate to the work.	1	0
MATHEMATICS: **Teacher's instruction promotes the following:**	Present	Not Present
Use of mathematical tools (e.g., technology, rulers, protractors) in ways that support mathematical exploration.	1	0
Use of mathematical tools (e.g., technology, rulers, protractors) in ways that enable students to solve mathematics problems.	1	0
Students' consideration of the strengths and limitations of a mathematical tool (e.g., technology, rulers, protractors) when students are deciding to use it in their work.	1	0
ELA-L: **Teacher's instruction promotes the following:**	Present	Not Present
Use of technology tools and digital media for local or global communication and collaboration.	1	0
Use of technology tools and digital media to conduct research.	1	0
Students' strategic choices of technology tools and digital media as an integrated component of ELA-L projects.	1	0

4. Facilitate Student Discourse

This practice emphasizes instruction that provides opportunities for students to engage in substantive interaction and dialogue with one another.

Example: A district rubric states that teachers should encourage students to "explain their thinking" to peers.

Facilitate Student Discourse		
GENERAL:		
Teacher provides opportunities for students to engage in classroom discourse that requires them to do the following:	Present	Not Present
Illustrate their thinking by communicating it not only to the teacher, but also to their peers.	1	0
Justify answers using evidence in classroom discussions.	1	0
Analyze other students' thinking.	1	0
MATHEMATICS:		
Teacher provides opportunities for students to engage in classroom discourse that requires students to do the following:	Present	Not Present
Explain their mathematical thinking and approaches to solving mathematics problems.	1	0
Justify explanations or provide evidence for answers in classroom discussions.	1	0
Critique the mathematical thinking of others.	1	0
ELA-L:		
Teacher provides opportunities for students to engage in classroom discourse that requires students to do the following:	Present	Not Present
Work on small-group, collaborative tasks that require illustrating thinking and building students' speaking, listening, and language skills, including collaborating in a productive and respectful manner.	1	0
Justify explanations or provide evidence for answers in classroom discussions.	1	0
Evaluate a speaker's point of view, reasoning, and use of evidence and rhetoric.	1	0

5. Engage in Purposeful Questioning

This practice stresses instructional approaches that use questioning to advance students' reasoning, deepen their understanding, further develop ideas, and strengthen their use of evidence.

Engage in Purposeful Questioning		
GENERAL: **Teacher uses questioning to do the following:**	Present	Not Present
Advance student reasoning and understanding (i.e., through higher order questioning).	1	0
Encourage students to explain their thinking	1	0
Encourage (could be require) students to justify their claims, using appropriate evidence.	1	0
MATHEMATICS: **Teacher uses questioning to do the following:**	Present	Not Present
Engage students in making sense of the "why" as well as the "how" in mathematics.	1	0
Encourage students to explain their thinking about mathematics problems.	1	0
Encourage students to use mathematics to justify their explanations and reasoning.	1	0
ELA-L: **Teacher uses questioning to do the following:**	Present	Not Present
Require students to read closely to advance understanding of texts.	1	0
Encourage students to explain their interpretations or understandings of texts.	1	0
Require students to use evidence from the text to justify increasingly complex within-text and cross-text analyses.	1	0

6. Integrate Key Components of Content and Dispositions of the Domain

This practice emphasizes instructional approaches that integrate a subject's key components, attending to multiple, discipline-specific competencies and dispositions (or habits of mind) in a lesson or unit.

Integrate Key Components of the Domain		
GENERAL: **Teacher's instruction does the following:**	Present	Not Present
Attends to and integrates key components of the domain, addressing content as well as the habits of mind, proficiencies, and dispositions within the discipline.	1	0

MATHEMATICS: **Teacher's instruction attends to and integrates the following:**	Present	Not Present
Both procedural fluency and conceptual understanding in mathematics, including the "why" as well as the "how" in mathematics.	1	0
Mathematical content and practices/processes.	1	0
ELA-L: **Teacher's instruction attends to and integrates the following:**	Present	Not Present
21st century ELA-L competencies, including at least two of the following: communication and collaboration in local and global contexts, understanding of other perspectives and cultures, comprehension and critique of a range of literary and informational text, strategic use of digital media and technology, argumentation and substantiation, and research (e.g., learning from and writing to texts).	1	0

7. Promote the Use of Academic and Domain-Specific Language and Vocabulary

This practice emphasizes instructional approaches that specifically and intentionally model and encourage the development and appropriate use of language specific to the domain, including formal vocabulary appropriate to an academic classroom setting.

Promote the Use of Academic and Domain-Specific Language and Vocabulary		
GENERAL: **Teacher's instruction does the following:**	Present	Not Present
Asks students to use academic and domain-specific vocabulary.	1	0
Connects students' informal language use with formal language use.	1	0
Demonstrates and draws attention to correct use of academic and domain-specific vocabulary.	1	0
MATHEMATICS: **Teacher's instruction does the following:**	Present	Not Present
Asks students to use precise mathematical language or notation.	1	0
Connects students' informal mathematical vocabulary or symbols with formal vocabulary or symbols.	1	0
Models use of appropriate or precise mathematical language or notation.	1	0

ELA-L: Teacher's instruction does the following:	Present	Not Present
Asks students to use academic language across the four ELA-L strands: reading, writing, speaking and listening, and language.	1	0
Models and facilitates students' use of newly acquired vocabulary and standard English usage.	1	0
Emphasizes Tier 2 and Tier 3 vocabulary by providing definitions and examples of academic and domain-specific vocabulary, as well as access to multiple print and digital sources of definitions.	1	0
Connects students' informal language usage and standard English.	1	0

8. Assess Important Skills and Understandings

The practice emphasizes teachers' use of assessment to track students' progress in mastering the concepts important to the academic domain.

Assess Important Skills and Understandings		
GENERAL: **Teacher uses assessment to do the following:**	Present	Not Present
Determine student progress in developing the important skills and understandings characteristic to the domain.	1	0
MATHEMATICS: **Teacher uses assessment to do the following:**		
Assess students' procedural fluency and conceptual understanding in mathematics.	1	0
Assess students' ability to engage with the mathematical practices (e.g., problem solving, reasoning).	1	0
ELA-L: **Teacher uses assessment to do the following:**	Present	Not Present
Assess the degree to which each student has met the ELA-L learning goals for content and concepts.	1	0
Assess students' ELA-L strategies and skills in reading (e.g., close and critical reading), writing, speaking and listening, and language (e.g., grammar use and vocabulary acquisition).	1	0

B. Instructional Practices Supporting Common Core–Aligned Standards, Specific to Mathematics

Embed Mathematics in Real-World Contexts and Emphasize Modeling

Embed Mathematics in Real-World Contexts and Emphasize Modeling		
MATHEMATICS: **Teacher's instruction does the following:**	Present	Not Present
Embeds mathematics in real-world contexts as appropriate.	1	0
Emphasizes the use of mathematics to model real-world phenomena, where appropriate.	1	0

Support Students' Productive Struggle in Mathematics

Support Students' Productive Struggle in Mathematics		
MATHEMATICS: **Teacher supports students' productive struggle in mathematics in the following ways:**	Present	Not Present
Provides opportunities for students to engage with challenging mathematics material.	1	0
Supports students as they grapple with challenging material without immediately "doing it for them."	1	0

Promote the Use of Multiple Representations and Connections Among Them

Promote the Use of Multiple Representations and Connections Among Them		
MATHEMATICS: **Teacher's instruction does the following:**	Present	Not Present
Emphasizes use of mathematical representations.	1	0
Makes explicit the ways in which different mathematical representations relate to each other.	1	0

Encourage Abstract Reasoning

Encourage Abstract Reasoning		
MATHEMATICS: **Teacher's instruction does the following:**	Present	Not Present
Encourages students to represent problem-solving situations with mathematics symbols, graphs, and pictures.	1	0
Fosters students' use of mathematical representations (symbolic and other) to solve the mathematics problem.	1	0
Encourages students to interpret mathematical solutions in the context of the problems posed.	1	0

C. Instructional Practices Supporting Common Core–Aligned Standards, Specific to English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects

Support Early Reading, Foundational Skills (K–5 ELA-L Teachers Only)

Support Early Reading Foundational Skills (K–5 Teachers Only)		
ELA-L: **Teacher's instruction does the following:**	Present	Not Present
Within reading lessons, provides explicit and systematic instruction focused on grade-level print concepts, phonological awareness, phonics, and fluency.	1	0

Teach Writing for Research, Argumentation, and Narrative (K–12 ELA-L and 6–12 Content Area Teachers)

Teach Writing for Research, Argumentation, and Narrative		
ELA-L: **Teacher's instruction does the following:**	Present	Not Present
Offers explicit writing instruction, focusing on writing to and from sources.	1	0
Assigns grade-appropriate and subject-appropriate writing tasks, including opinion, narrative, argument, and research.	1	0
Asks students to demonstrate writing processes such as prewriting, drafting, and revising.	1	0
Asks students to write for extended and shorter time frames and for various tasks, audiences, and purposes.	1	0

Build Independence for All Students: Providing a K–12 Literacy Staircase to the College and Career Readiness Anchor Standards (K–12 ELA-L and 6–12 Content Area Teachers)

Build Independence for All Students: Providing a Staircase to the College and Career Readiness Anchor Standards		
ELA-L: **Teacher's instruction does the following:**	Present	Not Present
Supports diverse student populations, including English learners, by providing all students with a range of high-quality grade- and subject-appropriate texts so that all students comprehend texts at the high end of the College and Career Readiness text complexity band independently and proficiently by the end of Grade 12.	1	0
Provides multiple opportunities for student discussion and peer review.	1	0
Provides ongoing specific feedback on multiple individual assignments and projects.	1	0
Incorporates multiple instructional strategies (e.g., scaffolding, variety of learning modes and styles) to support all students reaching independence and proficiency in grade- and subject-specific literacy tasks by the end of Grade 12.	1	0

Appendix B:
Rubric Scores on Common Practices

Exhibit B1. Summary of State and District Rubrics and Their Subject-Specific Indicators[a]

State Control	State	District	Rubric Based on Charlotte Danielson FFT	Number of Standards[b]	Number of Indicators[c]	Subject-Specific Components
Low	CA	CA CSTP	N	6	38	No subject-specific standards, domains, or other sections of rubric.
Moderate	CO	Durango	N	8	10	No subject-specific standards, domains, or other sections of rubric.
		Garfield	N	4	13	No subject-specific standards, domains, or other sections of rubric.
		Silverton	N	3	17	Standard 1 has elements specifically targeting literacy and writing.
		Thompson	N	3	14	Standard 3 has an element that mentions "student and literacy development and math."
		LEAP	N	4	13	No subject-specific standards, domains, or other sections of rubric.
		State Model	N	5	27	Standard 1 says that a teacher must demonstrate mastery in his or her subject area: literacy, mathematics, science, social studies, arts, physical education, or world languages. Standard 1 references specific ELA-L and mathematics concepts to be demonstrated by all teachers. Standard 1 also has sections specifically for elementary teachers teaching language arts or reading; secondary teachers teaching English, language arts, or reading; and teachers teaching mathematics.
Low	FL	Brevard	Y	5	25	No subject-specific standards, domains, or other sections of rubric.
		Broward	N	9	41	No subject-specific standards, domains, or other sections of rubric.
		Pinellas	N	5	33	No subject-specific standards, domains, or other sections of rubric.

State Control	State	District	Rubric Based on Charlotte Danielson FFT	Number of Standards[b]	Number of Indicators[c]	Subject-Specific Components
Low	KY	State Model	Y	4	22	No subject-specific standards, domains, or other sections of rubric.
		Boyd	Y	4	22	No subject-specific standards, domains, or other sections of rubric.
		Clinton	Y	2	10	"Explanation/use of oral/written language."
		Fayette	Y	4	21	No subject-specific standards, domains, or other sections of rubric.
		Floyd	Y	13	15	No subject-specific standards, domains, or other sections of rubric.
		Henderson	N	5	5	Standard 5 references "research-based and literacy strategy implemented."
		McLean	N	6	66	No subject-specific standards, domains, or other sections of rubric.
		Mercer	N	2	2	No subject-specific standards, domains, or other sections of rubric.
		Nelson	N	4	4	No subject-specific standards, domains, or other sections of rubric.
		Pike	Y	9	9	No subject-specific standards, domains, or other sections of rubric.
		Russell	Y	2	10	No subject-specific standards, domains, or other sections of rubric.
		Russell Independent	Y	2	10	Minor reference to ELA: "Use of Oral or Written Language."
		Trimble	Y	8	38	No subject-specific standards, domains, or other sections of rubric.
		Wayne	Y	5	23	Indicator in Standard 5 references "historical fiction titles."
Moderate	MA	State Model (MA Framework)	N	4	15	No subject-specific standards, domains, or other sections of rubric.
		CREST	N	9	8	No subject-specific standards, domains, or other sections of rubric.
		Easton	N	6	60	No subject-specific standards, domains, or other sections of rubric.

State Control	State	District	Rubric Based on Charlotte Danielson FFT	Number of Standards[b]	Number of Indicators[c]	Subject-Specific Components
High	NC	State Model (NC)	N	5	25	Indicator 3a references "Begins to integrate literacy instruction into selected lessons, Integrates effective literacy instruction throughout the curriculum and Incorporates a wide variety of literacy skills within content areas to enhance learning."
High	TN	Memphis	N	4	16	No subject-specific standards, domains, or other sections of rubric.
		State Model (TEAM)	N	3	19	No subject-specific standards, domains, or other sections of rubric.
n/a	CH	Achievement First	N	10	42	Indicator in Standard 5 references "analysis of a text."
		Alliance College-Ready	N	4	17	No subject-specific standards, domains, or other sections of rubric.
		KIPP	N	5	26	Indicator 4.2 Content Knowledge uses a mathematics example for "Has knowledge of what comes before and after in the curriculum." Indicator 4.3 Literacy for Everyone discusses reading and writing activities and tasks.
		STRIVE	N	4	2	Indicator in Standard 4 references student literacy.
		YES Prep	N	7	38	No subject-specific standards, domains, or other sections of rubric.
n/a	NAT	Danielson FFT Clusters- ELA	Y	6	31	All standards are subject specific.
		Danielson FFT Clusters-Math	Y	6	31	All standards are subject specific.
		SAP Lower ELA	N	4	17	All standards are subject specific.
		SAP Lower Math	N	3	15	All standards are subject specific.
		SAP Upper ELA	N	3	12	All standards are subject specific.

State Control	State	District	Rubric Based on Charlotte Danielson FFT	Number of Standards[b]	Number of Indicators[c]	Subject-Specific Components
		SAP Upper Math	N	3	15	All standards are subject specific.
		SREB ELA	N	4	9	All standards are subject specific.
		SREB Math	N	4	6	All standards are subject specific.
		TNTP	N	4	4	All standards are subject specific.

Note. FFT = Framework for Teaching; n/a = not applicable.

[a] Several rubrics, including those that do not include subject-specific standards or indicators, make some general acknowledgment of distinctions between subjects, saying, for example, that teachers should use "subject-appropriate" strategies or provide instruction on topics "important within the discipline."

[b] Standards also are called domains in some states.

[c] Indicators also are called components or elements in some states.

Table B2. Summary of Organization Rubrics and Their Subject-Specific Indicators

District	Number of Standards	Number of Indicators
Charlotte Danielson's FFT ELA-L Clusters	6	31
Charlotte Danielson's FFT Mathematics Clusters	6	33
SAP Lower ELA-L	3	13
SAP Upper ELA-L	3	12
SAP Lower Mathematics	3	13
SAP Upper Mathematics	3	15
SREB ELA-L	4	10
SREB Mathematics	4	6
TNTP	4	4

Table B3. State, Charter, and National Rubric Scores on Eight Common Practices

State	District	Establish Appropriate Learning Goals	Engage Students With Cognitively Challenging Tasks and Activities	Promote Strategic and Appropriate Use of Domain-Specific Tools and Resource	Facilitate Student Discourse	Engage in Purposeful Questioning	Integrate Key Components of the Domain	Promote Use of Academic and Domain-Specific Language and Vocabulary	Assess Important Skills and Understandings
NAT	Danielson FFT Clusters-ELA-L	0.67	0.25	0	1	0.5	1	0.25	0.5
	Danielson FFT Clusters-ELA-L	0.67	0.25	0	1	0.5	1	0.25	0.5
	Danielson FFT Clusters-Math	0.67	0	0	0.67	0.33	1	0.33	0
	Danielson FFT Clusters-Math	0.67	0	0	0.67	0.33	1	0.33	0
	SAP Lower ELA-L	0.33	0.5	0	0.33	1	0.5	0	0.5
	SAP Lower Math	0.67	0.5	0.33	0.67	0.67	0.5	0.67	0
	SAP Upper ELA-L	0	0.25	0	0.33	0.67	0	0	0.5
	SAP Upper Math	0.67	0.5	0.33	0.67	0.67	0.5	0.67	0
	SREB ELA-L	0.33	0.5	0	0	0.67	0	0	0.5
	SREB Math	0	0.75	0	1	0	0.5	0.33	0
	TNTP	1	0.33	0	1	0.67	0	0	1
CH	Achievement First	0.5	0.67	0	0.33	0.33	0	0.33	0
	Alliance College-Ready	1	0.67	0	0.67	0.67	0	0.33	1
	KIPP	1	0	0	0.67	0.67	0	0	1
	STRIVE	0	0	0	0	0	0	0	0
	YES Prep	1	0	0	0	0.33	0	0	1
CA	CSTP	0.50	0	0.50	0	0.33	0	0	1

State	District	Establish Appropriate Learning Goals	Engage Students With Cognitively Challenging Tasks and Activities	Promote Strategic and Appropriate Use of Domain-Specific Tools and Resource	Facilitate Student Discourse	Engage in Purposeful Questioning	Integrate Key Components of the Domain	Promote Use of Academic and Domain-Specific Language and Vocabulary	Assess Important Skills and Understandings
CO	Durango	0	0.67	0	0	0.67	0	0	0
	Garfield	0	0	0	0	0	0	0	0
	Jefferson	0	0.33	0	0	0	0	0	0
	LEAP	1	1	1	1	1	0	1	0
	Silverton	0	0.33	0	0	0	0	0	0
	State Model (CO State Model)	1	0.33	0	0	0	1	0	1
	Thompson	0.50	0	0	0	0	0	0.33	0
FL	Brevard	0	0.33	0	0	0	0	0.33	1
	Broward	0	1	0	0.33	1	0	0	0
	Pinellas	0.50	0.67	0.50	0	0.33	0	0	1
KY	Boyd	0	0	0	0	0	0	0	0
	Clinton	0	0	0	0	0	0	0	0
	Fayette	0	0.67	0.50	0.33	0.33	0	0	1
	Floyd	0	0.67	0.50	0.33	0	0	0	1
	Henderson	0.50	0	0	0	0.33	0	0	0
	McLean	0.50	0	0	0	0.33	1	0	0
	Mercer	0	0	0	0	0	0	0	0
	Nelson	0.50	0	0	0	0	0	0	1
	Pike	0.50	0	0.50	0	0.33	0	0	0
	Russell County	0	0.33	0	0	0	0	0	1

State	District	Establish Appropriate Learning Goals	Engage Students With Cognitively Challenging Tasks and Activities	Promote Strategic and Appropriate Use of Domain-Specific Tools and Resource	Facilitate Student Discourse	Engage in Purposeful Questioning	Integrate Key Components of the Domain	Promote Use of Academic and Domain-Specific Language and Vocabulary	Assess Important Skills and Understandings
KY	Russell Independent	0	0	0	0	0	0	0	0
	State Model (KY Framework)	0	0.67	0.5	0.33	0.33	0	0	1
	Trimble	0	0	0	0	0.33	0	0	0
	Wayne	0	0.67	0	0.33	0.33	0	0	0
MA	CREST	0.50	0	0	0	0	0	0	1
	Easton	0	0	0.50	0	0	0	0	0
	State Model (MA Framework)	0.50	0	0	0	0	0	0	1
NC	State Model (NC)	0.50	0.33	0.50	0	0	0	0	1
TN	Memphis	1	0.67	0	0	0.33	0	0	1
	State Model (TEAM)	0.50	0.67	0	0.33	1	0	0	1

Note. FFT = Framework for Teaching; NAT = national; CH = charter.

[a] Alignment scores for national rubrics and subject-specific rubrics include general, ELA-L, and mathematics elements.

Appendix C:
Scoring Methods and Interrater Reliability

The project team used the same scoring methods for all three waves of scoring. To score the rubrics, English language arts and literacy (ELA-L) and mathematics team members evaluated each standard, indicator, or descriptor (hereafter generally referred to as indicators) on a teacher evaluation rubric to identify points of alignment with each element on the American Institutes for Research (AIR) Alignment Protocol (AIR Protocol) using a matrix. If the teacher evaluation tool aligned with an AIR element, scorers indicated this alignment with a score of 1. The scorers gave indicators on the teacher evaluation document that did not align to the protocol a score of 0. Credit for alignment was given regardless of how many of the rubric's indicators aligned to each AIR element; a rubric with one descriptor that aligned to a certain element on the protocol received the same credit as a rubric with several descriptors aligned to that element. The total score for each element on the protocol was either 1 or 0. Total alignment for each teacher evaluation rubric was calculated for the general elements of AIR's Common Practices, for the combined general and subject-specific elements of the Common Practices, and for the subject-specific practices.

The project team used a multistep process to determine each state's alignment score. First, members of each team—ELA-L and mathematics—collaborated to calibrate their scoring and establish reliability. Recalibration meetings were held after each of the three waves of scoring. One member of each subject team was assigned to each general rubric; two coders from the same subject area were assigned to each subject-specific rubric.

The organizational structure of rubrics examined varied. Some grouped many teacher behaviors into an indicator, like the Danielson Framework for Teaching used by many Kentucky districts, while others had separate indicators for each teacher behavior, like Achievement First and Broward County, Florida. One-to-one alignment between the elements of the AIR Protocol and elements of a district or state rubric was determined to be inadvisable. Consequently, the decision was made that a "rich" indicator in a district or state rubric might be aligned to more than one AIR Protocol element. Conversely, in some cases multiple indicators in a rubric were coded to an AIR Protocol element when the combined set of indicators led to alignment with the element of instruction.

Computing Alignment Scores

The team calculated four types of *alignment scores* for each rubric:

- Common alignment score: Average of the eight Common Practice scores, accounting for the 16 general elements only.
- Common subject alignment scores: Average of the eight Common Practice scores, measuring alignment with all 16 general elements, 21 subject-specific elements in mathematics, and 23 subject-specific elements in ELA-L. The team computed scores separately among the eight Common Practices in mathematics and ELA-L.

- Specific alignment scores: Averages of the four mathematics practice scores and of the three ELA-L practice scores.
- Total alignment scores: Computed separately for mathematics and ELA-L—for mathematics, the average of all eight general practice scores and four mathematics-specific practice scores; for ELA-L, the average of all eight general practice scores and three ELA-L–specific practice scores.

The purpose of calculating multiple scores for each rubric was to take a more nuanced view of alignment. In writing the AIR Protocol, the team assumed that teacher evaluation rubrics that were fully aligned to practices supporting the Common Core State Standards would be subject specific. However, the team also assumed that most states and districts would have uniform rubrics and would not yet have instruments that met this standard. This study, then, serves as an opportunity to offer feedback to districts in achieving greater coherence and to offer a comparison between two approaches: unified rubrics and subject-specific instruments. Exhibit C1 displays the alignment scores and their constituent elements.[5]

Exhibit C1. Definitions of Scores Calculated

	Common Practices (8)				
Alignment Score	General Elements	Mathematics-Specific Elements	ELA-L–Specific Elements	Mathematics Practices	ELA-L Practices
Common alignment	X				
Mathematics common alignment	X	X			
ELA-L common alignment	X		X		
Mathematics-specific alignment				X	
ELA-L–specific alignment					X
Mathematics total alignment	X	X		X	
ELA-L total alignment	X		X		X

[5] See Welch, Potemski, et al.'s (2015) report for details about how each alignment score was calculated.

Interrater Reliability

Each general rubric was scored by an ELA-L coder and a mathematics coder (two coders, each from the same content area, coded the subject-specific rubrics). All scores were then reviewed by two additional content experts, who served as auditors. The ELA-L and mathematics auditors are content experts who worked on the original state-level alignment scoring project in spring 2015. Auditors scored three rubrics in preparation for the review and met with both teams to compare scores and to establish final consensus ratings for these instruments.

In this project, scoring happened in three waves, with team meetings after each wave. There was a meeting in mid-October for alignment analysis in subject groups and on December 4, 2015, for the scoring group as a whole. Based on these meetings, slight edits were made to the elements of the AIR Protocol to improve precision of intent and to the rules for coding to ensure consistency in ways of handling some of the new rubric features encountered. Final reliability checks were held by content area in mid- to late May and July 2016.

Interrater reliability was measured in terms of initial agreement on coding done by individual content experts. The team measured agreement as the number of identically coded cells on all rubrics. All teams eventually came to 100% agreement on the three rubrics chosen for the reliability check and made subsequent edits to the other instruments they scored individually. Reliability scores were as follows:

- Initial reliability for the team overall was 86% (or disagreement on alignment among rubric indicators and AIR Protocol elements in 14% of the cases).
- The mathematics team's initial reliability was 91%, and the ELA-L team's initial reliability was 75%.

Most of the disagreement among coders concerned three of the common instructional practices: Common Practice 2 (Cognitive Demand), Common Practice 5 (Questioning), and Common Practice 8 (Assessment). Initial reliability on the other five Common Practices was greater than 80%. These lower levels of initial agreement appeared in scores reported on Common Practices 2, 5, and 8 by both ELA-L and mathematics teams. Discussions in the three reliability audits helped bring the eventual scores within the subject teams to 100%.

In addition to coming to an agreement on all initial scoring differences, the teams have revised scoring rules and wording on the AIR Protocol to strengthen initial reliability in scoring future rubrics, focusing on Common Practices 2, 5, and 8. An anchor document with sample "borderline" indicators and alignment decisions also was developed to document team decisions and inform future scoring. This document is incomplete, mainly because of the lack of aligned examples for some elements.

One common measure of interrater reliability is percentage agreement among pairs of scorers. Thirty-seven district, state, or charter general teacher evaluation rubrics were coded. The average percentage of interrater agreement was 86% across all rubrics. Initial scorer agreement ranged from 50% to 100% by rubric and from 75.7% to 97.3% by element, although all rubrics eventually came to 100% agreement after consultation. The distributions of interrater agreement by rubric and by element are shown in Exhibit C2 and Exhibit C3.

Exhibit C2. Number of Rubrics at Different Levels of Interrater Agreement

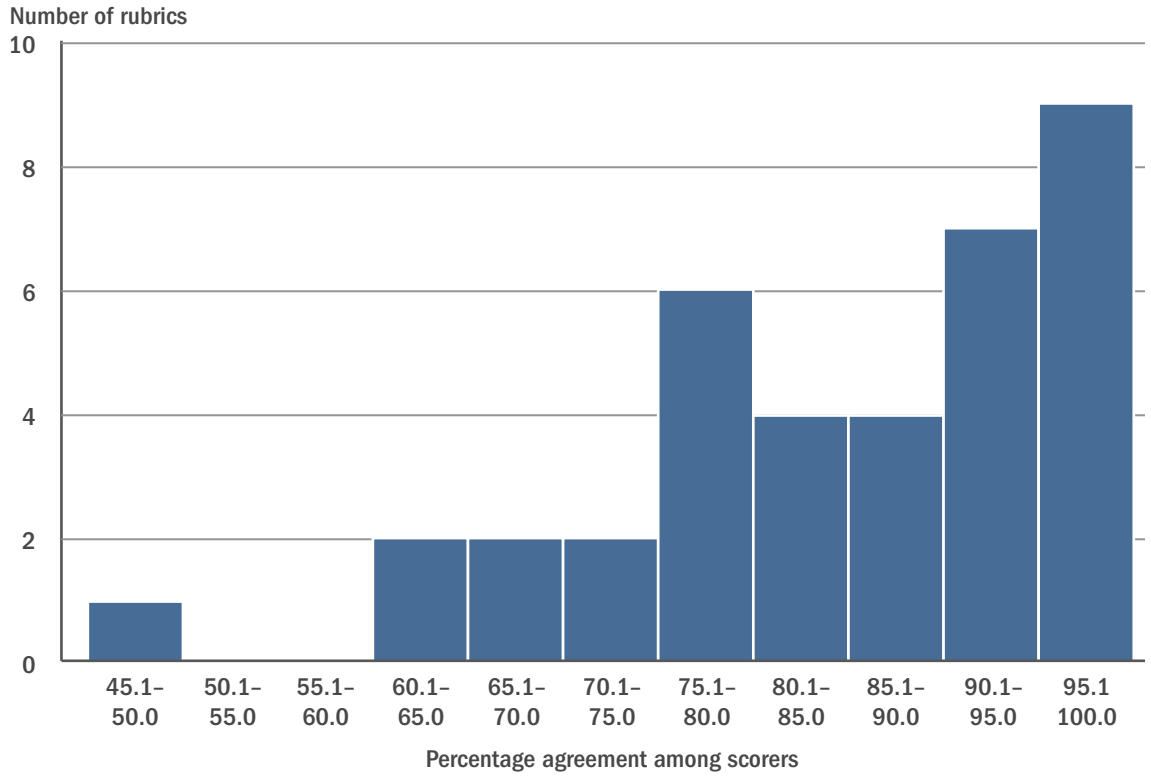
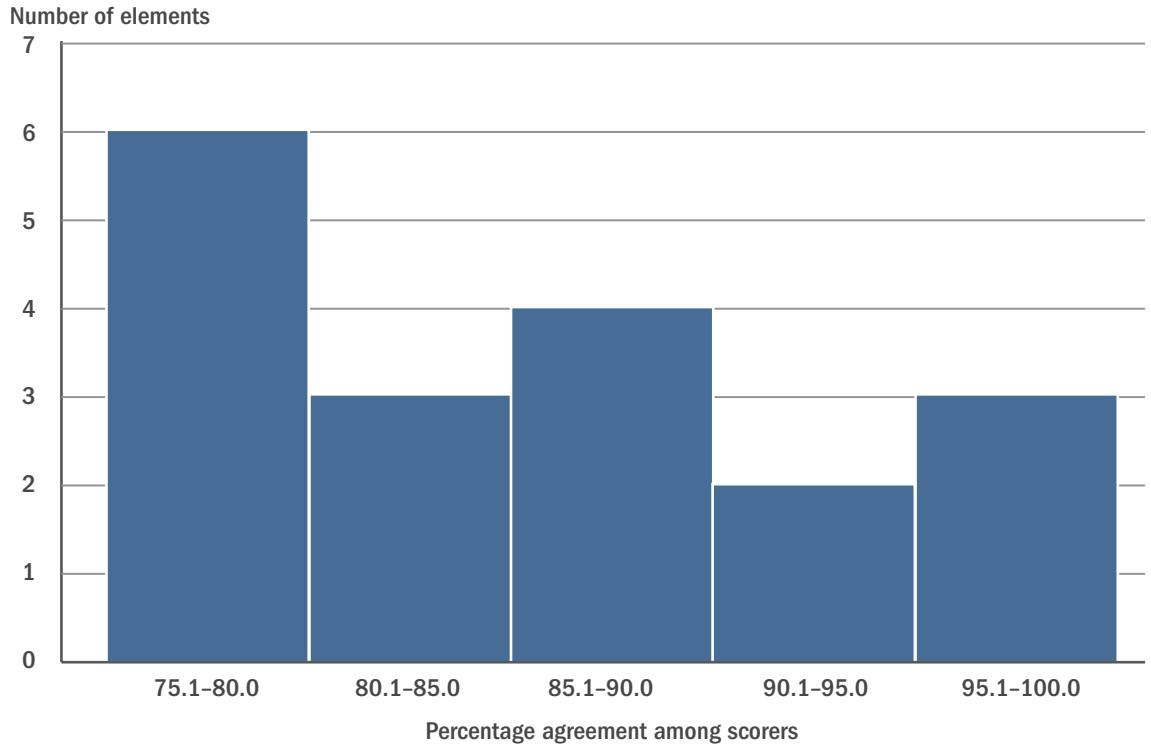


Exhibit C3. Number of Elements at Different Levels of Interrater Reliability



Although commonly used, interrater agreement is not the best measure of reliability of scoring because it does not take into account chance rater agreements. Cohen's kappa coefficient is a statistic designed to measure interrater reliability for categorical variables, taking chance agreement into account. The decision was also made to use Cohen's kappa to measure interrater agreement based on the belief that this more robust measure will provide a better picture of AIR Protocol performance and the confidence that can be placed in scoring results. Kappa values ranged from 0.00 to 0.874. Although there are no universal standards for describing the strength of interrater agreement based on kappa vales, 0.80 and up is considered strong interrater agreement, values of 0.60–0.79 indicate moderate agreement, values of 0.40–0.59 indicate weak agreement, values of 0.21–0.39 indicate minimal agreement, and values of 0.00–0.20 indicate no agreement. Scoring for six elements showed strong agreement, six showed moderate agreement, four showed minimal agreement, and two showed no agreement. Results by element are provided in Exhibit C4.

The large difference between kappa values and overall percentage agreement is accounted for by the general lack of indicators in the evaluation rubrics to the elements of the AIR Protocol. Scorers were mostly in agreement when alignment to an element was not present, but the general, nonspecific wording in many teacher evaluation rubrics makes it possible, but not necessary, that an element could be aligned to the indicator. This uncertainty reduced reliability of scoring as scorers tried to interpret the meanings in teacher evaluation rubrics. The elements with the lowest kappa scores were those seen very rarely, and the high percentage of agreement that these elements were not aligned was not matched on the few occasions when alignment was possible.

Exhibit C4. Values for Cohen's Kappa, Sensitivity, and Specificity of Agreement by Element

Element	Practice	Element Statement	Kappa Value	Sensitivity	Specificity
1a	Establish Appropriate Unit Learning Goals	Align to the state standards for students at that grade or level.	0.784	81%	80%
1b	Establish Appropriate Unit Learning Goals	Represent a coherent progression of content, skills, and materials or resources over time.	0.427	40%	71%
2a	Engage Students With Cognitively Challenging Tasks and Activities	Think deeply about the material (e.g., synthesize information/ concepts), reason about the content, or make connections across topics.	0.504	55%	65%
2b	Engage Students With Cognitively Challenging Tasks and Activities	Engage with or apply their understanding and skills to challenging or complex tasks or texts.	0.501	55%	89%
2c	Engage Students With Cognitively Challenging Tasks and Activities	Engage in critical and evaluative analysis.	0.262	20%	75%

Element	Practice	Element Statement	Kappa Value	Sensitivity	Specificity
3a	Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources	Strategic use of tools or resources, including technology, to support work in the domain.	0.513	50%	72%
3b	Promote Strategic and Appropriate Use of Domain-Specific Tools and Resources	Student involvement in a decision-making process through which to determine which tool is appropriate to the work.	0.000	0%	95%
4a	Facilitate Student Discourse	Illustrate their thinking by communicating it, to not only the teacher but also their peers.	0.455	44%	91%
4b	Facilitate Student Discourse	Justify answers using evidence in classroom discussions.	0.680	57%	97%
4c	Facilitate Student Discourse	Analyze other students' thinking.	0.874	80%	69%
5a	Engage in Purposeful Questioning	Advance student reasoning and understanding (i.e., through higher order questioning).	0.622	68%	68%
5b	Engage in Purposeful Questioning	Encourage students to explain their thinking.	0.368	36%	72%
5c	Engage in Purposeful Questioning	Encourage (or require) students to justify their claims, using appropriate evidence.	0.306	25%	83%
6	Integrate Key Components of the Domain	Attend to and integrate key components, proficiencies, and dispositions within the discipline.	0.303	20%	89%
7a	Promote Use of Academic and Domain-Specific Language and Vocabulary	Ask students to use academic and domain-specific vocabulary.	0.843	75%	97%
7b	Promote Use of Academic and Domain-Specific Language and Vocabulary	Connect students' informal language use with formal language use.	0.000	0%	97%
7c	Promote Use of Academic and Domain-Specific Language and Vocabulary	Demonstrate and draw attention to correct use of academic and domain-specific vocabulary.	0.444	33%	89%
8	Assess Important Skills and Understandings	Determine student progress in developing the important skills and understandings characteristic to the domain.	0.675	71%	73%

Appendix D: Protocol Development

The project team began developing the American Institutes for Research (AIR) Alignment Protocol (AIR Protocol) by identifying instructional practices intended to promote learning aligned with the Common Core State Standards in mathematics and English language arts and literacy (ELA-L). We identified these aligned instructional practices by reviewing eight practice frameworks designed by leading educational organizations that have written about the instructional shifts they see as associated with the adoption of new standards. The frameworks reviewed were as follows:

- AIR's Center on Great Teachers and Leaders Core Instructional Practices (Leo & Coggshall, 2013)
- Insight Core Framework (2013)
- Interstate Teacher Assessment and Support Consortium (InTASC) Model Core Teaching Standards (2013)
- National Council of Teachers of Mathematics Principals to Action: Mathematics Teaching Practices (2013)
- Student Achievement Partners Instructional Practice Guides (2013a, 2013b, 2013c, 2013d, 2013e)
- Teaching Works High Leverage Practices (2012)
- TNTP Core Teaching Rubric (2014)
- Tri-State ELA and Mathematics Rubrics (2013a, 2013b, 2013c)

The team used these frameworks to create two comprehensive lists of instructional practices, one focused on mathematics and one on ELA-L. Next, the team condensed these lists, grouping together common and related practices represented in multiple sources. The team also referred to the original Common Core State Standards themselves in mathematics and ELA-L to supplement practices appearing in the frameworks. The team found some parallel practices supporting the implementation of the mathematics and ELA-L standards, while some practices were unique to one of these two practice areas.

The result was a list of eight Common Practices that support the adoption of Common Core–aligned standards in mathematics and ELA-L; four instructional practices that support the adoption of mathematics standards; and three practices that support the adoption of ELA-L standards. These practices comprise the Alignment Protocol found in Appendix A.

AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW | Washington, DC 20007-3835 | 202.403.5000 | **www.air.org**

6918_10/16