Appendix F

HANDBOOK FOR IMPLEMENTING PACT SCORER TRAINING

March, 2006

Table of Contents

Introduction

The purpose of this Handbook is to provide guidance for implementing PACT training. The Handbook was designed for two primary audiences: 1) program directors and other individuals assigned responsibility for implementing the PACT scoring training; and 2) trainers of scorers. It represents our best thinking at this time, based on experience, both within and outside PACT. Logistical constraints will undoubtedly require some adaptations. When compromises are necessary, please think carefully about how to achieve the purposes of the training, which are: 1) to acquaint scorers with the scoring system and help them understand the "1", "2", and "3" scoring levels for each Guiding Question; and 2) to enable scorers to consistently judge candidate performance in relationship to these standards. This is necessary to ensure the reliability and validity of the scoring process.

In Spring 2006, benchmarks are available in the areas of Elementary Literacy, Elementary Mathematics, History/Social Science, Mathematics, and Science. Pre-scored Teaching Events, but not benchmarks, will be available in English-Language Arts. No benchmarks are available for Art, Music, Physical Education, and World Languages. Those scorers will need to be trained in another content area and transfer the principles learned to their own content areas.

Page 1 and pages 48 through 54 describe arrangements needed to enable the training to take place, options for implementing the training, potential pitfalls, and resource materials from the PACT website. These pages will be of interest to anyone with responsibility for implementing some part of the training.

Pages 2 through 47 contain detailed descriptions of the training activities and the training materials, aside from the Teaching Event Handbook, rubrics, benchmarks, benchmark analyses, and scoring forms. These pages will be of most interest to trainers and those who wish to become better acquainted with the training process for scoring the PACT Teaching Event.

If you have any questions or uncertainties about conducting training, please e-mail Kendyll Stansbury at kendylls@stanford.edu and/or call her at (650) 724-9262.

Recruitment

This section includes descriptions of the desired characteristics of scorers and trainers of scorers. It is the responsibility of Program Directors to either identify these educators or to assign responsibility to others for identifying them.

Identifying Scorers

Desired characteristics:

- an understanding of the content and the variety of pedagogical approaches to teaching the content for the credential associated with the Teaching Event
- familiarity with beginning teaching and common challenges for beginners
- the ability to evaluate the merit of teaching approaches that differ from their own
- experience in evaluating teachers in a variety of instructional and community settings

Sources of scorers include:

- Faculty
- Supervisors
- Cooperating teachers
- Other teachers and school administrators
- National Board certified teachers or candidates for National Board certification
- BTSA staff or support providers
- Content specialists or staff developers in local districts
- Staff or teachers associated with Subject Matter Projects

Identifying Trainers

Trainers need to have the same characteristics of scorers, plus:

- Previous experience scoring Teaching Events, with evidence of reliable scoring
- Previous experience as a trainer
- Substantial knowledge of how to match evidence to rubrics
- Good judgment about when to be flexible and when to be firm
- Knowledge of the Teaching Performance Expectations (TPEs)

Training Design and Materials

This section includes a detailed description and approximate timeline for training activities, and training materials aside from the Teaching Event Handbook, rubrics, benchmarks, benchmark analyses, and scoring forms. The latter are distributed to either the Program Director or directly to the trainers by the Stanford PACT staff. It is the responsibility of the trainers to study the training protocol and benchmarks prior to training, double check that all materials are duplicated prior to the training, and to carry out the training activities.

RECOMMENDED TRAINING AGENDA

To make the reading go more quickly, we recommend that scorers should read the Thinking Behind the Rubrics, the Level 2 benchmark, and its completed scoring form prior to Day 1.

Day 1

9:00 Introduction

9:30 Orientation to Scoring Process in Subject-Specific Groups

10:30 Begin review of Level 2 benchmark

12:15 Lunch

1:00 Continue review of Level 2 benchmark

4:45 Adjourn

Day 2

9:00 Comparison of Levels 1 and 2

11:30 Lunch

12:15 Comparison of Levels 2 and 3

3:00 Calibration Teaching Event & Debriefing

5:00 Adjourn

Training Protocol for Two Day Training

Day 1

9:00 Welcome, Introductions, Housekeeping, Training Schedule

- Display Slide 1 (PACT Scoring Training) as participants are arriving. Welcome the participants to the training. Tell them that the purpose of the training is to help them understand the scoring system well enough to reliably score Teaching Events.

- Introduce yourself. Ask participants to introduce themselves, modeling a brief introduction with relevant information, e.g., I am ___ and I teach the methods course" or "My name is _____; I supervise student teachers at A, B, and C Elementary Schools.

- After introductions are completed, then go through housekeeping details, e.g., where the restrooms are, food. Point to the Questions/Parking Lot chart. Tell participants that if they have questions that are "off-topic" at the time, please write them on a Post-It and put them on the Questions/Parking Lot chart. The questions will be addressed at the end of the day. Although it is tempting to talk about suggestions for supporting students and program elements that are strong or weak relative to what is asked of candidates in the Teaching Event, these issues should be written on Post-its and exiled to the Parking Lot chart. Otherwise, it is unlikely that the training will be completed on time. During lunch and on breaks, the trainer can try to take time to group related questions/issues and maybe even respond briefly.

- Display Slides 2 (Day 1 Training Schedule) and 3 (Day 2 Training Schedule) and go over the training schedule. First, participants will be oriented to the scoring process, the match between the Teaching Event evidence and the TPEs, and the documentation of the ratings that they will turn in. Next, participants will think about the personal biases that we bring to scoring (as we all do) so that they can be mindful of them. Then, we will go over a Teaching Event that was scored a "2" for many Guiding Questions. We will try to understand what the "2" level means. Then we will divide into two groups. A homework assignment will be to skim the two remaining benchmarks, while taking notes and scoring three tasks across the two benchmarks. In discussing the evidence on Day 2, we will try to sharpen our knowledge of what the "1", "2", and "3" levels mean for each Guiding Question.

- Display Slide 4 (Goals) and ask participants to read the goals for the training. Explain that in order to score reliably, scorers must understand the likely sources of evidence for each Guiding Question, how to identify and record unbiased evidence, and how to match evidence to the rubric level descriptors.

Module 1

9:15 The Scoring Process

- Display Slide 5 (Structure of the Teaching Event). Points to make include:
  - ✓ Note the Instructional Context task, which is not scored. (The purpose is to help scorer understand about student needs and characteristics, available resources, and district/school expectations about teaching practice.)
  - ✓ Note that evidence for Academic Language is gathered across all tasks. Feedback from the first year was that there is not sufficient evidence to merit task-based rubrics addressing Academic Language, but that evidence across all tasks is sufficient to score it separately.
  - ✓ There are very specific directions for completing each section of the Teaching Event. Call the participant's attention to the Teaching Event prompts. Note that not all candidates follow the instructions or provide all the information requested. If candidates provide evidence in another part of the Teaching Event that is relevant to a different task, we note the evidence and judge whether it merits changing the score for any previously scored Guiding Questions. Other than that, we score what the candidate provides and do not try to infer what the candidate might have been thinking or intending based on limited information. Be cautious about overinterpretations or making large inferences. This is especially true when you are familiar with the curriculum materials, when the candidate provides little information, or when the information provided isn't making sense to you.

- Display Slide 6 (Task by Task Scoring). Explain the task-based scoring process. Note that the Teaching Event is not likely to be an exact match to a rubric level descriptor. We score by selecting the level which is the closest match, the one for which there is a preponderance of evidence.

- Emphasize that we want scorers to apply the scoring rubrics as written. As we go through the benchmark Teaching Events, we will be explaining the intent behind each rubric. The goal of this training is to help them understand the score points 1, 2, and 3 for each Guiding Question. These will be the performances that are most common. Any suggested changes to the rubrics go on the Parking Lot.

- Call their attention to the second page of the scoring form, titled Confidence in Ratings.

  - ✓ The Confidence in Ratings scale gives a scorer the opportunity to look at the profile of ratings across different aspects of teaching and report the degree of confidence they have in the rating profile from the scoring rubrics. This data gives us evidence about the scorers' perceptions of the validity of the ratings. Low levels of confidence raise a red flag that requires attention.
  - ✓ Call their attention to the second question on that page: Based on the evidence in the Teaching Event, what is your holistic impression of this candidate? Scorers should use their professional judgment independently of

our scoring rubrics to form this impression. This will also help us check on the validity of our scoring system.
  - ✓ There is also a place where scorers can note if there is something unusual about this Teaching Event that either might affect the validity of the scores.
  - ✓ The next question that asks whether you know the candidate whose Teaching Event you are scoring and if so, in what capacity. There is some anecdotal data to suggest that candidates are scored higher by those who know them; this question provides systematic data to see if that is true.
  - ✓ Lastly, if the candidate's scores are fairly steady across tasks and there are no complications affecting scoring (e.g., the candidate's writing is difficult to interpret), please recommend it as a potential benchmark. This cuts down on the number of Teaching Events we need to read to find benchmarks.

- Ask participants to take out the handout, "Relationship of TPEs to Guiding Question Rubrics." The TPEs are the standards for this assessment. Note that all TPEs are assessed through the Teaching Event, some more strongly than others. However, not all aspects of all TPEs are assessed; that is beyond the scope of any single assessment.

- State that the development teams in each content area used the K-12 Student Academic Content Standards and Curriculum Frameworks to inform the selection of important foci, so there is a connection to the K-12 curriculum as well.

9:30 Bias

- This is a potentially sticky subject, so try to keep a light tone. Here are points to make:
  - ✓ We all carry biases with us. This does not meant that we are bad people, just that we are heavily influenced by our preferences and experiences.
  - ✓ Some biases are based on research and are appropriate reflections of our professional knowledge. Other biases stem from our preferred theoretical perspective and our limited exposure to different instructional practices, teaching contexts, and types of students. Still other biases (positive and negative) are absorbed unconsciously through messages from media, friends, and relatives.
  - ✓ We all have preferences about the type of teaching we like. We wouldn't be human if we didn't. We also can better evaluate practices with which we are very familiar in contexts with which we are familiar. Even research-based biases can be difficult to apply fairly to an unfamiliar context.
  - ✓ Nonetheless, fairness in scoring dictates that we try hard to set aside the biases that are not grounded in research and try to apply fairly those that are grounded in research.

- Ask scorers to take out and silently read the handout "Notes about Bias" to identify possible sources of bias. After five minutes, check to see if everyone is done. Ask for one or two comments on what they read or experiences with bias to share.

- On either a blank overhead or a piece of chart paper, draw a T-chart with one side headed "Teaching Practices I Like" and the other side headed "Teaching Practices I Dislike". Begin by sharing a personal example of each.

- State that the same situation may trigger different types of bias in different people. For example, one person may feel that a candidate teaching a class composed mostly of English learners has a more difficult placement than other candidates, and try to judge this performance more leniently. Another person might believe that English learners are in particular need of good instruction, and have a higher standard for instruction for these students than for a class of native English speakers. Ask scorers to take out the handout "Teaching Practices I Like/Dislike" and take five minutes to write up their own list of teaching practices that they like and dislike, to identify the type of bias (positive or negative) likely to be triggered.

- After a few minutes, ask if anyone is willing to share an example of their own preferences. Stress that we all have preferences; the task is to be aware of them so that they don't interfere with fair scoring. Try to get 3-4 examples of each. <u>Do not</u> debate whether the biases are appropriate, just write them down. Note that our preferences influence scoring. Ask scorers to be especially careful in identifying evidence and matching it to the rubric language when they are reviewing a Teaching Event that includes teaching practices that they especially like or dislike to guard against bias.

- Note that scorers also need to be mindful of their current emotional state. Fatigue or frustration will make it more difficult to be fair. This is a sign that you need to take a brief break, e.g., to the snack table or for a brief walk, to refresh yourself.

10:00 Break

10:10 Notetaking and Documentation

- Display Slide 7 (Guiding Questions and Rubrics). These are the Guiding Question categories. Each Guiding Question is scored by one rubric. We'll examine the rubrics in more depth throughout the training.

- Display Slide 8 (About the Rubrics) Points to include:
  - ✓ The rubrics were based on the professional knowledge and experience of the developers plus trends in Teaching Events from previous years.
  - ✓ Level 1 was constructed to reflect candidates with some skill but who need one more semester of student teaching before they are ready to be in charge of a classroom. These candidates do not meet performance standards. Since this is the lowest level, candidates with few discernable skills are placed here as well, but that is not how the level was defined.
  - ✓ Level 2 was constructed to reflect a judgment of "ready to be in charge of a classroom", but just adequate. For more challenging Guiding Questions, this

level reflects quite modest skills and abilities. The expectation is that candidates at Level 2 have a foundation of knowledge on which to build and will get better with more support and experience. At this level, the candidate demonstrates an acceptable level of performance on the standards.
- ✓ Level 3 represents candidates who have a solid foundation of knowledge and skills. These candidates demonstrate an advanced level of performance on the standards relative to most beginners.
- ✓ Level 4 represents the stellar candidates, the top 5 % or so of candidates. Because they are so rare, we have trouble identifying what they might look like. If you have a Teaching Event that does not meet the rubric criteria but seems to you to be an outstanding performance with respect to a particular Guiding Question, then score it using the rubric and write us a note on the second page of the scoring form recommending that we take a look at this Teaching Event for redefining Level 4. Be sure to indicate the relevant Guiding Question.

- Display Slide 9 (Scorer Work for Each Task). Describe what scorers are going to do as they score each Teaching Event. Make the following points:
  - ✓ Only list the key evidence that supports the rubric rating on the scoring form. Scoring forms are available electronically, at your program's option.
  - ✓ Another scorer should be able to read the description of evidence and patterns of evidence and see how it relates to the selected rubric level.
  - ✓ Performances can reflect aspects of different scoring levels. Select a scoring level by asking, "To which level do the practices documented most closely correspond? (This is a "preponderance of evidence" approach.) The benchmarks were selected to avoid borderline performances. If you have a borderline performance with respect to a particular Guiding Question, it will take more time to determine the preponderance of evidence.
  - ✓ Note that when scoring is completed, scores are transferred to the cover page for data entry.

- We are trying to figure out how to speed up notetaking and scoring without sacrificing quality. We discovered last year that scorers vary in their notetaking preferences, and want to accommodate effective practices. Scorers should feel free to take notes in the approach with which they are most comfortable. The Thinking Behind the Rubric document identifies the big ideas for each Guiding Questions as well as distinctions between rubric levels, and can be a valuable tool to guide notetaking.

- Display Slide 10 (Note-taking). Points to include:

  - ✓ Focus on notes that represent evidence most strongly related to rubric levels. As you become more familiar with the rubrics, you will be able to be more selective in taking notes.

- ✓ Instead of using an adjective/adverb, try to describe succinctly what about the evidence makes it "good", e.g., "Visuals used to convey meaning of concept". This will make the notes more objective and less judgmental.
- ✓ One clue that you aren't focusing on objective evidence is if you find yourself thinking something like "What s/he must have been doing was…", "I can't believe that s/he would have done ___ without doing _____". Thoughts along these lines should alert you to come back to what you know that the candidate actually said and did. We try to see a consistent story of teaching in these Teaching Events, but sometimes the Teaching Events are patchworks of thoughtful insights and omissions/confusions or uncritical borrowing of instructional strategies.
- ✓ The notes should capture aspects of the teaching performance that are related to concepts in the rubric.
- ✓ Evidence should be related to the language in the rubric. Point out an example on the first page of the scoring form for the first Benchmark.

- Display Slide 11 (Specificity of Notes). Go over the three examples to illustrate the desired specificity of notes. In addition, make the following points:
  - ✓ Scorers can take notes on post-its, on a pad of paper, or directly on each rubric page. Please don't take notes on or highlight the actual Teaching Event unless you are explicitly told by your program that you can do so. Some of the Teaching Events will be double scored, and the second scorer should not be influenced by the first scorer's notes or highlighting.
  - ✓ If post-its are not used to note evidence, then some brief description of where the evidence is found should be included. The reason for noting the source of evidence is that if scorers want to re-examine the evidence at any time, they can find it quickly. It doesn't matter how it is referenced – a page number, the name of the document, a rough location in a video clip.
  - ✓ In selecting a score, scorers have found it helpful to highlight the language in the rubric levels that corresponds to the patterns of evidence recorded. Language in more than one level may be highlighted, but only one level will be selected as reflecting the preponderance of evidence. What the highlighting seemed to do was to help the scorer focus on matching the evidence to the rubric. Portraying the matches visually helped in selecting the score. Feel free to use the highlighting if it seems helpful. One caution, however, is that the preponderance of evidence is not quantitative. You might have one phrase highlighted in one level and two phrases highlighted in another. The evidence supporting the one highlighted phrase might be so compelling that it trumps the evidence for the two phrases, making the level with only one highlighted phrase the better match for the preponderance of evidence.

- Display Slide 12: (Characteristics of the Recorded Summary of Evidence).
  - ✓ Record key pieces of evidence on the scoring form that justify the score. If appropriate, you can just move selected post-its or copy electronic excerpts from the Teaching Event to the appropriate page. (The benchmark write-ups

are overexplained for training purposes; you don't need that level of detail.) Our eventual goal is to figure out how to record key pieces of evidence on the scoring form to justify the rating without requiring extensive reorganization or rewriting of notes. Any suggestions on strategies for doing this are appreciated.

- ✓ Put more effort into documenting the scoring of 1 ratings or borderline performances between Levels 1 and 2 that are rated a 2. Those ratings are most likely to result in failing the Teaching Event when failing to meet a passing standard has consequences. (The plan is to have more than one scorer read a failing Teaching Event when it results in a high stakes decision.) Evidence for ratings of solid 2s through 4s is less critical to document for purposes other than candidate feedback.

- Display Slide 13 (Let's honor the credential students and their supporters). Explain that we are about to look at a completed Teaching Event, and you want to go over some caveats. Here are some points to include as you go over this slide:
  - ✓ Stress the importance of following professional norms about maintaining the confidentiality of a specific candidate's performance.
  - ✓ Remind participants that these are candidates who have spent less than a year in professional training in pedagogy. We need to be respectful when we discuss these performances with a trainer or when we compare ratings with someone else who scored the same Teaching Event. However, we also need to be critical of the evidence that we see. The expectations for novices are built into the rubrics, and we should judge the evidence accordingly. If scorers feel that the rubric expectations are too high, then they should give us feedback, but they should apply the rubrics as they are written.
  - ✓ As scorers, we need to avoid the mindset of a supervisor that tries to identify important areas for improvement, looking at what is not going well and thinking about how to best help the candidate. As a scorer, look at what the candidates <u>are</u> doing and not at what they are <u>not</u> doing. Evidence of what they <u>are</u> doing maps on the rubric; evidence of what they are <u>not</u> doing is only reflected in the rubric in extreme cases and results in the lowest rating.

- Ask for any questions at this point.

Module 2

10:30 Understanding the Instructional Context

- Ask scorers to take out their first benchmark and, if the group is large, form scorers into pairs or trios. Display Slide 14 (Reviewing the Description of the Context for Learning) and review. Stress that they are looking for relevant features of the context that are likely to impact candidates' teaching decisions. Remind participants of the focus of the learning segment. Give groups about ten minutes to read the Context for Learning task and identify relevant features. Then ask them to share out popcorn style and chart the information.

- Close the activity by reminding the scorers that we ask candidates to include a description of the Instructional Context so that we might better understand how they respond to their students and the advantages or constraints under which they are teaching. Remind scorers that we are expecting to see candidates striving for good teaching practice. Also caution them to be aware of possible biases related to the teaching context, either positive or negative, that they need to guard against as they read the rest of the Teaching Event.

10:45 Understanding Level "2": Planning

- Call their attention to "academic language" and any other term that needs to be clarified before they read the Planning Task. Talk about the distinctions between academic language and content understandings (acknowledging the difficulty in disentangling the two for Elementary Literacy, English/Language Arts, or Modern Languages).

- Display Slide 15 (Artifacts and Commentary). There are two sources of evidence in the Teaching Event: artifacts and commentaries. The content of the artifacts – lesson plans, video, student work- can be judged independently of the teacher's commentary. The commentaries provide explanations of decisions and sometimes additional contextual information about students, learning goals, or the curriculum materials. Sometimes the evidence from commentaries and artifacts support each other, as when the commentary draws your attention to some feature of their lesson plan that reflects tailoring to their students' particular needs that you might have otherwise overlooked or failed to understand. Sometimes the evidence from commentaries and artifacts contradict each other, e.g., when a candidate states that they are developing student abilities to think independently and provides instructions for completing learning tasks that involve very little thinking. Sometimes the commentaries reflect criticism of the instructional strategies and materials required to be used. Using the evidence in the artifacts as well as the evidence in the commentary, scorers can use their professional judgment to draw conclusions about the depth of the candidate's understanding of the teaching knowledge and skills addressed by the Guiding Question that is being scored.

- Display Slide 16 (Planning Instruction and Assessment - Evidence). Explain that this Task focuses on a candidate's *plans* as written. Since candidates typically get their plans from other sources, e.g., textbook companies, published curricula, other teachers, it is tricky to score the candidate's competence and not that of the source of the plans. You shouldn't make inferences about how the plan works; the commentary asks the candidate to explain how the plan builds student understanding, so score that and what the candidate says in the description of the plans. Note the evidence to be provided in this task, and call scorers' attention to the actual prompts in the Handbook. Contrast the evidence about assessment in the plans, i.e., the collection of assessments to monitor student learning during the learning segment vs. the evidence that will come later in the Assessment task, i.e., using evaluative criteria (or rubrics) to interpret evidence of student learning and using that evidence to inform future planning.

- At this point, you introduce the scorers to the rubrics for which they will be recording evidence from the Planning Task. Note that this includes both the Planning rubrics and the Academic Language rubrics.

  - ✓ Display Slide 17 (Planning Rubrics). Ask participants to take out a Scoring Form on which they will take notes about the rubrics. State that this Teaching Event will be used to illustrate the "2" level of the scoring rubrics. It is not always scored at a 2 for every Guiding Question. In those cases, we will go over the reasons why it was scored differently and talk about what a 2 performance might look like. For each Guiding Question, ask participants to identify the key concepts in Level 2 and artifacts and commentary prompts that are most likely to produce evidence. If you have any tips, cautions, or issues related to reading the Planning Task for the Level 2 benchmark, share them.
  - ✓ Call participants' attention to Level 3 of the third Planning rubric. The phrase "productive…and receptive…modalities" sometimes causes confusion. Productive modalities are when students are producing their own understanding. Receptive modalities are when students are communicating their understanding of something someone else produced. For example, in a discussion, a student who offers a new point is in a productive modality, but a student who *responds* to the point is communicating his/her understanding of the first student's point as well as possibly communicating his/her own understanding.
  - ✓ Display Slide 18 (Academic Language). Ask participants to read the "Thinking Behind the Rubric" discussion of the Academic Language rubric. Distinguish Academic Language (GQ10) from Accessible Content (GQ2), acknowledging that there is considerable overlap between reading, writing, speaking, and listening in the curriculum and understanding of the content. However, there are strategies, e.g., visual and other nonverbal representations of content, which are more effective at providing access to the curriculum than in developing academic language. Pairing English Learners with more fluent English speakers may or may not develop the academic language of the English Learner, but is more likely when norms and structures to guide interactions are provided. Scorers should think carefully about whether strategies appropriate for English Learners reflect access to content, development of academic language, or both. Explain that the Academic Language rubric will be scored with evidence gathered across tasks. It is the only rubric for which this is true, and has a different structure than the other rubrics.
  - ✓ Display Slike 19 (Academic Language rubrics). Point out the key concepts in the rubric for which evidence is likely to be gathered from the Planning Task: the description of student strengths and needs with respect to language development, how academic language is represented in the learning goals and

plans, how academic language understanding and use is scaffolded. If the notetaking strategy does not involve writing on the rubric, scorers need to think about how they can keep track of the evidence for Academic Language, since it can come from any task. Two options are a different color of post-it or some sort of symbol, like a star, that distinguishes these notes from other notes.

- Rubric by rubric, go over the evidence and how it maps to the rubric language for the Planning rubrics. Do this as a whole group, eliciting examples of evidence from scorers. Focus the discussion on characteristics of Level 2; if the GQ was scored at level 1 or 3, then emphasize the differences between Level 2 and the evidence. If the score was adjusted based on later evidence, point that out and describe the additional evidence and how it changed the scoring decision. If the evidence made it difficult to choose between levels, acknowledge that and describe what tipped the decision to a particular level. Stop at the end of each rubric for questions, but try to not get bogged down. Acknowledge any problematic rubrics, identifying the problems on a post-it on the Parking Lot chart.

- In the discussion, try to focus the scorers on what the candidate <u>was</u> doing, and not on what the candidate was failing to do. Discuss the pitfall of overgeneralizing of one piece of evidence, and the need to focus on patterns of evidence within the task. Remind scorers that we are using a preponderance of evidence to guide decisions.

12:15 Lunch

1:00 Understanding Level 2: Instruction

- Display Slide 20 (Instructing Students and Supporting Learning - Evidence). Go over the evidence available from the task. Talk about how this task gives us the only evidence where we see teachers actually interacting with students. This allows a well-informed judgment about the candidate's statements.

- Display Slide 21 (Instruction Rubrics). For each Guiding Question, ask scorers to identify the key concepts in Level 2 for each rubric and the commentary prompts which are most likely to provide evidence. (The video clips are likely to provide evidence for both Guiding Questions.) Sometimes watching the learning tasks being implemented helps scorers understand the learning tasks better. In that case, it is all right to go back and record additional evidence for the Planning rubrics and, if it is key evidence, adjust the score either up or down.

- There is both video and a commentary. If the evidence in the video contradicts the candidate's statements in the commentary, e.g, the candidate says that s/he is asking students questions to make them think and the candidate is doing all the talking in the video except for a few yes/no questions, then scorers should go with the evidence in the video. If you and your fellow Lead Trainers have any tips, cautions, or issues related to reading the Instruction Task, share them.

- Point out the key concepts in the Academic Language rubric for which evidence is likely to be gathered from the Instruction Task: scaffolds or supports for student to enable students to meet language demands of the learning tasks.

- Ask participants to read the summaries for the two Instruction rubrics. Ask them to watch for evidence supporting the summaries in the video clip(s). Ask them to also separately record evidence of language supports for the Academic Language rubric. Show the video clip(s) and then ask participants to skim the commentary.

- Rubric by rubric, go over the evidence and how it maps to the summaries and to the rubric language. Focus the discussion on characteristics of Level 2; if the GQ was scored at level 1 or 3, then emphasize the differences between Level 2 and the evidence. If the score was adjusted based on later evidence, point that out and describe the additional evidence and how it changed the scoring decision. If the evidence made it difficult to choose between levels, acknowledge that and describe what tipped your decision to a particular level. Stop at the end of each rubric for questions, but try to not get bogged down. Acknowledge any problematic rubrics, making appropriate notes on post-its.

2:00 Break

2:10 Understanding Level 2: Assessment

- Display Slide 22 (Assessing student learning - Evidence) and go over the key pieces of evidence. Tell scorers that the Assessment Task focuses on the interpretation of student work using evaluative criteria or rubrics, and the uses of those interpretations for drawing conclusions about student learning and for making future plans about instruction. State that scorers should read the student work samples so that they can judge the accuracy of the candidate's analysis and also note areas that the candidate may be missing. Display Slide 23 (Assessment rubrics). For each Guiding Question, ask scorers to identify the key concepts in Level 2 for each rubric and the artifacts and commentary prompts which are most likely to provide evidence. Call their attention to the Academic Language rubric, and ask participants to identify artifacts and commentary prompts that are likely to provide evidence about academic language development.

- Rubric by rubric, go over the evidence and how it maps to the summaries and to the rubric language. Focus the discussion on characteristics of Level 2; if the GQ was scored at level 1 or 3, then emphasize the differences between Level 2 and the evidence. If the score was adjusted based on later evidence, point that out and describe the additional evidence and how it changed the scoring decision. If the evidence made it difficult to choose between levels, acknowledge that and describe what tipped your decision to a particular level. Stop at the end of each rubric for questions, but try to not get bogged down. Acknowledge rubrics that are difficult to apply.

3:30 Understanding Level 2: Reflection

- Display Slide 24 (Reflecting on Teaching and Learning - Evidence) and go over the key pieces of evidence, including the daily reflections. Note that the Reflection Task focuses on what a candidate has noticed and learned from teaching the learning segment. Display Slide 25 (Reflection rubrics). For each Guiding Question, ask scores to identify the key concepts in Level 2 and the artifacts and commentary prompts most likely to provide evidence. Call participants' attention to the first Reflection rubric. Tell them that the evidence for this rubric focuses on adjustments to instruction that are actually made during the learning segment; if the adjustments are for the next time the learning segment is taught or if it is not likely that they were made, then the evidence maps to the second Reflection rubric.

- Rubric by rubric, go over the evidence and how it maps to the summaries and to the rubric language. Focus the discussion on characteristics of Level 2; if the GQ was scored at level 1 or 3, then emphasize the differences between Level 2 and the evidence. If the score was adjusted based on later evidence, point that out and describe the additional evidence and how it changed the scoring decision. If the evidence made it difficult to choose between levels, acknowledge that and describe what tipped your decision to a particular level. Stop at the end of each rubric for questions, but try to not get bogged down. Acknowledge rubrics that are difficult to apply.

4:00 Understanding Level 2: Academic language

- Go over the evidence collected for academic language development. Have participants gather the evidence that they previously recorded for the Academic Language rubrics. Ask participants to read the summary and discuss how the evidence supports both the summary statements and the assigned score. Try to use the discussion to illuminate characteristics of Level 2, regardless of how the Academic language rubric was scored.

4:30 Debrief

- Ask for questions, concerns about the scoring process and application of rubrics. Address questions on the Questions/Parking Lot chart, grouping similar questions. Suggest that if questions occur to scorers overnight that they post them before training resumes.

- Tell participants that they need to start at 9:00 sharp the next day. **4:45 Adjourn**

Day 2

Module 3

9:00 Understanding Level 1: Planning

- Welcome participants back. Remind participants of the Parking Lot, and tell them that questions that occur during the day will be responded to at the end of the day, if there is time, or by e-mail. Explain that today, we will focus on understanding Level 1 for each Guiding Question. We will not address Level 4 unless a particular Guiding Question was scored as Level 4. We will begin by paying particular attention to the distinctions between Levels 1 and 2, even when the Teaching Event was not scored at Level 1.

- Ask participants to review Level 1 and the differences between Level 1 and Level 2 for the three Planning rubrics and the Academic Language rubrics, then read and take notes on the Planning commentary.

- Rubric by rubric, go over the evidence and how it maps to the summaries and to the rubric language. Focus the discussion on how the characteristics of Level 1 and 2 differ; if the GQ was scored at level 2 or higher, then try to keep the discussion of how the evidence fits the score point short, and ask the group to brainstorm examples of a 1 performance that contrasts with the performance scored.

- If the score was adjusted based on later evidence, point that out and describe the additional evidence and how it changed the scoring decision. If the evidence made it difficult to choose between levels, acknowledge that and describe what tipped your decision to a particular level. Stop at the end of each rubric for questions, but try to not get bogged down. Acknowledge rubrics that are difficult to apply.

10:20 Break

10:30 Understanding Level 1: Instruction

- Ask participants to review Level 1 and the differences between Level 1 and Level 2 for the two Instruction rubrics, then read and take notes on the Instruction commentary. Remind them to take notes on evidence for Academic Language. Show the video clip(s).

- As you and the group discuss each rubric, briefly remind participants of the salient features of the "2" level, and ask them to read the distinctions between levels 1 and 2 from the Thinking Behind the Rubrics. Rubric by rubric, go over the evidence and how it maps to the summaries and to the rubric language. Focus the discussion on how the characteristics of Level 1 and 2 differ; if the GQ was scored at level 2 or higher, then try to keep the discussion of how the evidence fits the score point short, and ask the group to brainstorm examples of a 1 performance that contrasts with the performance scored.

11:00 Understanding Level 1: Assessment

- Ask participants to review Level 1 and the differences between Level 1 and Level 2 for the two Assessment rubrics, then read and take notes on the Assessment commentary. Remind them to take notes on evidence for Academic Language.

- Rubric by rubric, go over the evidence and how it maps to the summaries and to the rubric language. Focus the discussion on how the characteristics of Level 1 and 2 differ; if the GQ was scored at level 2 or higher, then try to keep the discussion of how the evidence fits the score point short, and ask the group to brainstorm examples of a 1 performance that contrasts with the performance scored.

11:30 Understanding Level 1: Reflection

- Ask participants to review Level 1 and the differences between Level 1 and Level 2 for the two Reflection rubrics, then read and take notes on the Reflection commentary. Remind them to take notes on evidence for Academic Language.

- Rubric by rubric, go over the evidence and how it maps to the summaries and to the rubric language. Focus the discussion on how the characteristics of Level 1 and 2 differ; if the GQ was scored at level 2 or higher, then try to keep the discussion of how the evidence fits the score point short, and ask the group to brainstorm examples of a 1 performance that contrasts with the performance scored.

12:00 Lunch

12:45 Understanding Level 1: Academic Language

- Briefly remind participants of the salient features of the "2" level, and ask them to read the distinctions between levels 1 and 2 from the Thinking Behind the Rubrics. Discuss the evidence for each benchmark with respect to academic language. Focus the discussion on how the characteristics of Level 1 and 2 differ; if the GQ was scored at level 2 or higher, then try to keep the discussion of how the evidence fits the score point short, and ask the group to brainstorm examples of a 1 performance that contrasts with the performance scored.

Module 4

1:15 Understanding Level 3

- There is not time to read a Level 3 benchmark in a two-day training. Give participants a brief overview of the context and focus of the Level 3 benchmark. Task by task, ask participants to read silently the differences between Levels 2 and 3 in *Thinking Behind the Rubrics*, and then the relevant parts of the scoring form, noting how the evidence maps to at least Level 3 (or what else it needs to map to Level 3, if it is a Level 2). Ask them to jot down any questions. Give them 45 minutes for this task.

- At 2:00, go through each task, making key points and then asking for questions. Here are points to make:
  - ✓ Planning rubrics: The clear connections emphasized in the first planning rubric should be evident in the candidate's description of the plans, and not inherent in the curriculum itself. Some candidates (and teachers) do not teach the curriculum as written, so scorers need to be alert to the way that the candidate is using the curriculum materials and not make big inferences from their knowledge of the curriculum. For the third planning rubric, remember that the reference to assessments showing "some depth of understanding or skill" is in the context of a 3-5 lesson sequence (elementary) or 3-5 hours of instruction (secondary). Think about what would be surface-level vs. deeper understandings developed in that length of time. They won't be as deep as understandings of big ideas that are developed across the year. Also keep this in mind for introductory learning segments and learning segments for primary grades.
  - ✓ Instruction rubrics: For Level 3 on the second rubric, if students are working in groups, the candidate may be listening to different groups and, if things are going well, not offering any response. Candidates should not be penalized if students are doing fine without any intervention.
  - ✓ Assessment rubrics: The distinctions between Levels 2 and 3 don't seem to cause any problems.
  - ✓ Reflection rubrics: For the first reflection rubric, adjustments to instruction at Level 3 can also include the kind of adjustments specified in Level 2 as long as there is also a pattern of adjustments focused on specific learning needs, either individual or collective. Some candidates are very thorough in their reflections, and they should not be penalized for considering a broad range of instructional improvements.
  - ✓ Academic language: For the second academic language rubric, there should be a high degree of specific language supports to earn a Level 3 rating that include the three specified components: explicit models, opportunities for practice, and feedback. If there are a variety of strategies sustained throughout the learning segment that contain the three specified components of language support, candidates who scored at Level 2 for the first academic language rubric can score at Level 3 for the second one.

2:45 Break

3:00 Calibration Teaching Event and Debriefing

- Distribute a Teaching Event from this year to be scored for calibration. (If one is not yet available or scoring will occur much later than scoring, this activity can be delayed until later.) Select a time to show the video clip(s) and announce when and where they will be shown.

- At 4:30, chart the distribution of scores and discuss any discrepancies.

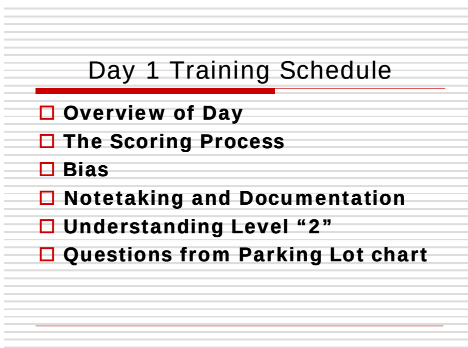
5:00 Adjourn

- Thank participants for their time and contributions. Tell them that when they score, they will have an opportunity to complete an on-line survey to provide comments, suggestions, and ideas to help improve the training and scoring process. If there is time, address questions on the Questions/Parking Lot chart; if not, assure them that answers will be forthcoming by e-mail.

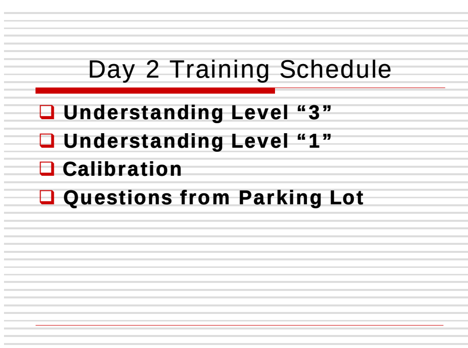
PowerPoint Slides

Slide 1

Slide 2



Slide 3



Slide 4

Goals

- Scorers will identify potential sources of bias in scoring.
- Scorers will understand the expected sources of evidence and TPEs assessed for each Guiding Question.
- Scorers will understand key features of the rubric levels, how to gather evidence, and how to match evidence to rubric levels.
- Scorers will understand the notetaking process and the level of documentation expected.
- Scorers will understand the distinctions between levels 1, 2, and 3 of the rubrics.

Slide 5

Structure of the Teaching Event

<table>
<tr><td colspan="4">Context for Learning
• Instructional Context form
• Context for Learning commentary</td></tr>
<tr><th>Planning</th><th>Instruction</th><th>Assessment</th><th>Reflection</th></tr>
<tr><td>• Daily Lesson Plans
• Handouts, overheads, etc.
• Planning Commentary</td><td>• Video clip(s)
• Instruction commentary</td><td>• Analysis of whole class achievement + 3 student work samples
• Analysis of learning for two of the three students
• Assessment commentary</td><td>• Daily reflections
• Reflective commentary</td></tr>
<tr><td colspan="4">Academic Language
(Evidence gathered across tasks)</td></tr>
</table>

Slide 6

Task by Task Scoring

- Read Task 1 for Context. Score Task 2 (Planning) completely, then move on to 3
- Score Task 3 (Instruction) completely (consulting Task 2 evidence as needed)
- Score Task 4 (Assessment) completely (consulting Tasks 2 & 3 evidence as needed)
- Score Task 5 (Reflection) completely (consulting Tasks 2, 3, & 4 evidence as needed)
- Score Academic Language rubrics, consulting evidence collected across the five tasks.

Slide 7

Guiding Questions and Rubrics

PLANNING
- Establishing a Balanced Instructional Focus
- Making Content Accessible
- Designing Assessments

INSTRUCTION
- Engaging Students in Learning
- Monitoring Student Learning During Instruction

ASSESSMENT
- Analyzing Student Work From an Assessment
- Using Assessment to Inform Teaching

REFLECTION
- Monitoring Student Progress
- Reflecting on Teaching

ACADEMIC LANGUAGE
- Understanding language demands
- Supporting academic language development

Slide 8

About the Rubrics

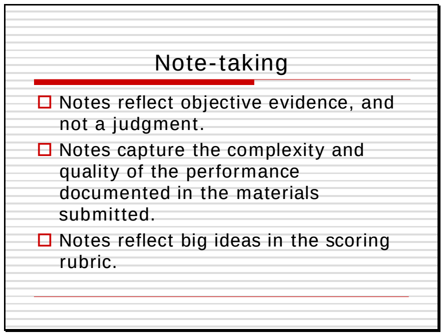
- Grounded in the notion of a continuum of teacher learning, practice, & expertise
- "2" is deemed to be an acceptable level for the preservice teacher
- Need to honor the performance level of 2 for a student teacher

Slide 9

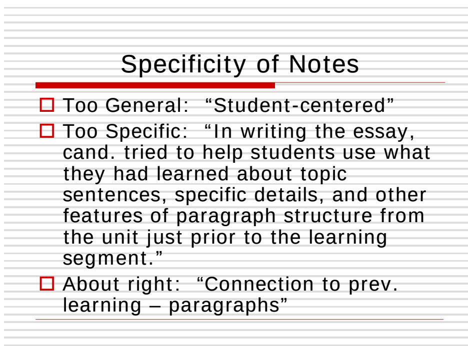
Scorer Work for Each Task

- While reading, take notes on evidence related to the big ideas from the rubrics for that task
- After reading a task, review the notes. On each rubric page, copy or summarize key features of evidence and patterns from the notes
- Assign the relevant score for the rubric. If in doubt, consult the *Thinking Behind the Rubrics*.

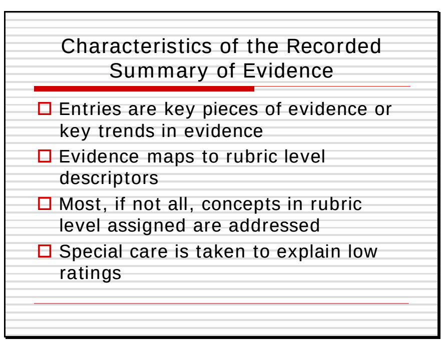
Slide 10



Slide 11



Slide 12



Slide 13

Honor the Credential Candidates and the Educators Supporting Them

- Confidentiality about early career candidates who are allowing us to look in on their work
- These are novices; look for what they can do, not what they cannot yet do
- Respectful treatment of the students and our teacher education colleagues, as we peer in at these performances
- That said, this is a pilot test from which we can learn, so let's be (respectfully) sharp and critical so we all can make the system excel in what it captures, differentiates

Slide 14

Reviewing the Description of the Context for Learning

- Read Task A (Context for Learning), taking notes on features that might affect a candidate's teaching practice (5 minutes)
- In your group, share the features noted and make a list. (5 minutes)

Slide 15

Artifacts and Commentaries

- Artifacts provide independent evidence of how strategies are implemented.
- Commentaries provide windows into a teacher's thinking.
- Commentaries often provide additional contextual information.
- Artifacts and commentaries can be consistent or contradictory.

Slide 16

Planning Task - Evidence

- **Daily Lesson Plans**
- **Instructional Materials**: overheads, handouts, assessments, etc.
- **Planning commentary**: importance of focus to student learning, how plans build student understanding of focus, link of plans to students, language demands of tasks, use of assessment information, how special needs of students are addressed

Slide 17

Planning Rubrics

- Establishing a balanced instructional focus
- Making content accessible
- Designing assessments

Slide 18

Academic Language

- Addresses language development, not content understandings unrelated to use of language, e.g., conceptual understanding
- To score above Level 2, candidates need to provide supports or scaffolds for language use and comprehension, not just for content understandings.
- To score at higher levels, candidates need to go beyond teaching new vocabulary to address features of texts or oral/written tasks

Slide 19

Adademic Language Rubrics

- Understanding language demands
- Supporting academic language development

Slide 20

Instruction Task - Evidence

- **Video**: short required clip(s)
- **Instruction commentary**: up to 3 pp., routines and other structures, student engagement, addressing individual needs, language supports, evaluating success of strategies

Slide 21

Instruction Rubrics

- Engaging students in learning
- Monitoring student learning during instruction

Slide 22

Assessing Student Learning - Evidence

- **Evaluative criteria or rubric:** Criteria used to assess student work
- **Student work samples:** samples of student work to illustrate analysis
- **Assessment commentary:** Summary of class performance on criteria, analysis of student work, deeper analysis of two students, next steps in instruction

Slide 23

Assessment Rubrics

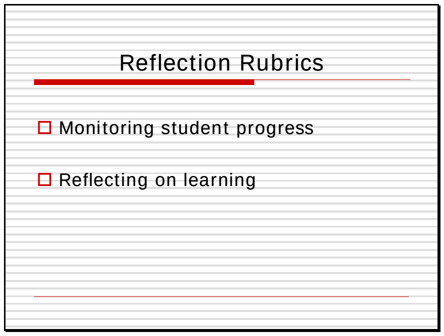
- Analyzing student work from an assessment
- Using assessment to inform teaching

Slide 24

Reflecting on Teaching and Learning - Evidence

- **Daily Reflections** – what's working, what's not, for whom and why, and implications for the next lesson
- **Reflective Commentary** – reflection on student learning results, what was learned about students as learners, changes to make if could do over

Slide 25


Reflection Rubrics
Monitoring student progress
Reflecting on learning

Relationship of TPEs to Guiding Question Rubrics
2005-2006 Pilot

Scoring Rubric	Teaching Performance Expectations (TPEs) Measured by Guiding Question Rubrics												
	TPE 1. Specific Pedagogical Skills for Subject Matter Instruction	TPE 2. Monitoring Student Learning During Instruction	TPE 3. Interpretation and Use of Assessments	TPE 4. Making Content Accessible	TPE 5. Student Engagement	TPE 6. Developmentally Appropriate Teaching Practices	TPE 7. Teaching English Learners	TPE 8. Learning about Students	TPE 9. Instructional Planning	TPE 10. Instructional Time	TPE 11. Social Environment	TPE 12. Professional, Legal, and Ethical Obligations	TPE 13. Professional Growth
GQ1: Balanced instructional focus													
GQ2: Accessible content													
GQ3: Assessment design													
GQ4: Engagement in learning													
GQ5: Monitoring learning													
GQ6: Analysis of student work													
GQ7: Assessment informing teaching													
GQ8: Monitoring student progress													
GQ9: Reflection on learning													
GQ10: Understanding language demands GQ11: Supporting academic language development													

Bias Review Handouts

Notes about Bias when Scoring Teaching Events

This document draws from a document prepared by Michal Lomask of the Connecticut State Department of Education and accounts of the bias training by the National Board for Professional Teaching Standards.

Scoring biases include any subconscious and/or conscious influence that may affect the scoring of the Teaching Event that is unrelated to the rubrics. We need to guard against letting our experiences and our preferences blind us to what is actually present in a teacher's performance. This results in drift from the established scoring criteria.

Here are some common scoring biases to avoid:

Source of Bias	Possible Effect on Scoring
Quality of writing: • Errors in grammar, spelling, and syntax • Fluent writing that uses jargon or language from the Teaching Event and/or that tells a story that you want to hear • Errors in logic • Use of non-standard forms of English	Candidates will differ in their writing ability and the ease with which they communicate. Scorers should not confuse the quality of writing or the use of phrases in the prompts with the quality of teaching, but instead look at details within the writing and the documentation provided to understand the teaching being described. Be especially wary when you are tired or when the current Teaching Event is an extreme contrast to the one just read. Also beware of nonstandard writing that may trigger race, class, or regional biases.
Technical quality of materials submitted: • Clear vs. haphazard organization of responses to prompts • Impressive use of graphic design or technological features • Quality of videography	Scorers need to take care to not be distracted by technical features of the Teaching Event that are extraneous to teaching. While they may be impressed at or distressed with a candidate's command of technology or organizational ability, this is irrelevant to scoring the Teaching Event.
Quality of instructional materials: • Commercial vs. teacher-made materials	Candidates who use materials from a strong commercial curriculum design may be unfairly credited with good instructional design, while candidates who design their own materials may suffer from the reverse effect.
Filling in gaps in descriptions of practice: • Lack of detail • Missing pieces	Candidates do not always follow instructions carefully, leaving gaps that scorers are tempted to fill. Be wary of large inferences. Inferring learning goals from materials is probably a low-level inference, but assumptions about

Source of Bias	Possible Effect on Scoring
	prior steps or a candidate's undescribed intentions are not warranted.
Grade level or teaching assignment	Grade levels or teaching assignments may be associated with higher or lower content expectations without considering the needs of the students in the class. Evaluate whether the teaching takes into account the prior achievement of students and language proficiency. This does not mean watering down the content, but does mean that a different approach may be needed to make the content accessible to students. Conversely, the instruction should also be appropriately challenging, particularly for students whose prior achievement suggests that they either have learned the content or can learn new content easily.
Emotional reactions: • To curricular materials • To the context of teaching • To teaching practice that you know you could improve with a few simple additions or adjustments • To the writing tone	Scorers may be disappointed when candidates do not teach as they would under the same circumstances. They also may be upset with teaching that is not as polished as they are used to. Try to understand the candidate's intent and the impact of the teaching on students, and to recognize what the candidate making a lot of unfortunate decisions is doing well. Also be alert to writing tones that predispose you to like or dislike a candidate and try to focus on describing what candidates are actually doing and how it is impacting students.
Halo effect:	Sometimes candidates do one thing spectacularly well or poorly, and this influences the rest of the ratings, even when unrelated. Some influences, e.g., poor match of materials to teaching goal, are difficult to disentangle from later decisions. Try to record evidence that is closely matched to the rubric language and judge it independently of previous evidence.
Leniency/Stringency	Some scorers may interpret the rubric language quite literally and others may give candidates a great bit of leeway in applying the scoring levels. This results in scores that are consistently higher or lower than the benchmark performances,

Teaching Practices I Like	Teaching Practices I Dislike

THINKING BEHIND RUBRICS
2005-06 Pilot

To help you apply the rubrics, the following describes some of the thinking behind the rubrics as they were being developed. The selection of a rubric level is based on a preponderance of evidence matching a rubric level description for that score point. The description requires professional judgment to apply to the evidence; it is not in the form of an item whose presence or absence is readily apparent to noneducators, and perhaps even to nonspecialists.

This document discusses each rubric in two sections: 1) Big ideas and their progression across rubrics and 2) Level differences that describe key differences between adjacent rubric levels.

Planning Rubrics

Many candidates use plans and materials developed by others – textbook writers, curriculum developers, the cooperating teacher, educators posting materials on the Internet, etc. This reflects teaching as a collaborative enterprise, which is desirable, especially for novices. However, it complicates scoring in determining to what extent the content and structure of the plans, instructional materials, and assessments reflect the candidate's own knowledge and skills. Therefore, for this task, the plans and other materials must be judged in light of the candidate's explanations of how the instructional and assessment materials work. Scorers should rely heavily on these explanations and not make big inferences from their knowledge of the curriculum intent or knowledge of how the curriculum might be optimally implemented. For example, if the mathematics curriculum is designed to develop students' conceptual understanding, but the candidate talks consistently in procedural terms, the candidate is probably attending only to procedural fluency. Candidates are prompted for explanations in the commentary but sometimes provide them in other parts of the Teaching Event, e.g., in the Instruction commentary or daily reflections.

1 Establishing a balanced instructional focus

Big ideas and their progression across the rubric:

- One vs. multi-dimensional central focus – Learning segments are defined by a central instructional focus. To have a central focus, the standards/objectives, learning tasks, and assessments are related to an identifiable theme, essential question, or topic within the curriculum. Across the rubric, this idea progresses from no discernable central focus or a one-dimensional focus at Level 1 to a multi-dimensional focus at Levels 3 and 4. ("Dimensions" are explained in the next bullet.)
- Connections between different types of knowledge in the content area – In every content area, the different types of knowledge constitute dimensions of the curriculum. These include more basic types of knowledge (e.g., facts, skills, conventions) and higher order knowledge or thinking skills (such as strategies for interpreting or reasoning from facts or evidence, synthesizing ideas, strategies for

evaluating work). Examples that define some subject-specific dimensions are specified in the rubric for each content area. Connections between the different types of knowledge are emphasized, since any one type in isolation provides a limited view of content understandings. Across the rubrics, this idea progresses from a uni-dimensional focus on one type of knowledge to the exclusion of other types to a design where the types of knowledge are not only present but consistently connected.

- Progression of learning tasks – A progression of learning and assessment tasks is where the tasks are structured and sequenced to build and check student understanding of the central focus. The progression may be linear, with each lesson providing prerequisite knowledge or skills for the next, or nonlinear, with each lesson's learning task contributing a different perspective on the central focus and together building deeper student understanding.
- Deep understandings of the central focus – Understandings that go beyond the superficial. This must be considered in light of the developmental level of the students, the short length of the learning segment, and whether the focus is being introduced to students or whether they have worked on it previously.

Level differences:

- Between 1 and 2: There are two ways **to achieve a Level 1 rating**: (1) The first is to lack any identifiable focus for the learning segment (e.g., a lack of even a general connection among the standards/ objectives, learning tasks, and assessments). **In all levels beyond Level 1**, the learning segment has a central focus. (2) The second way is through a learning segment that is defined by a number of lessons rather than a central focus and therefore contains one or more lessons that do not contribute to the central focus identified by the candidate.

    Elementary Literacy Only: In Elementary Literacy, a lack of focus can also be achieved through a learning segment that is centered on integrated instruction, but which does not focus on *literacy* learning. These learning segments typically focus on the application of previously learned reading and writing skills/strategies in the context of the content area, but lack instruction that is directed at moving the students beyond their current levels of literacy skill or understanding. For integrated instruction to be rated higher than a 1 on this rubric, it must include an explicit focus on literacy skills/strategies, such as new general literacy skills/strategies applied to subject-specific text(s) or specific characteristics of a text type to comprehend or compose subject-specific text.

  If there is a central focus, the contrast between Levels 1 and 2 is in the dimensionality of the focus. **At Level 1**, the standards/objectives, learning tasks and assessments focus exclusively on one type of knowledge to the exclusion of any others. **At Level 2**, one type of knowledge is very dominant, with only superficial, fleeting, or inconsistent attention given to other types of knowledge.
- Between 2 and 3: **At Level 3**, there are clear connections between the various types of knowledge. "Clear" means that the connections go beyond the superficial. This is true for either the instructional tasks or the assessment tasks, but not both. One type

of knowledge may be prominent, as it is the major focus of the learning objectives (i.e., the idea is not an equal balance between types). In that case, other types of knowledge will be used to either give meaning to or strengthen understanding of the prominent type. Another characteristic of **Level 3** is that the daily set of standards/objectives, learning tasks, and assessments work together to build a progressive understanding of the content. This progression can either be linear, where each lesson builds on the previous one, or nonlinear, where a concept, phenomenon, etc. is examined from multiple perspectives to build a more holistic and/or nuanced understanding.

- Between 3 and 4: **At Level 4**, the connections happen for both the learning tasks *and* the assessment tasks. In addition, the progression provides students opportunities to deepen their understanding of the central focus of the learning segment.

2 Making Content Accessible

Big ideas and their progression across the rubric:

- Significant content inaccuracies – This only appears at Level 1. (If the content is inaccurate, it doesn't matter how it is taught.) Significance is defined in terms of impact on student learning. Examples of significant inaccuracies are an incorrect definition or conceptualization that carries across the learning segment or a consistent pattern of errors across the learning segment. The errors should be significant enough that the students are going to be disadvantaged when they encounter the same content or apply the same skills again.
- Relationship of students' experiential backgrounds, interests, or prior learning to the standards/objectives – This progresses from little or no relationship at Level 1 to strategic use of particular aspects of students' experiential backgrounds or interests as well as prior learning to help students reach the standards/objectives at Levels 3 and 4.
- Forms of student support – These include any strategy aimed at helping students who need additional support to understand the content and/or learn the skills inherent in the standards/objectives. A few examples of such strategies are listed in a footnote for the rubric. This idea progresses from general support for usually struggling students at Level 2 to strategies that work well together to address varied student learning needs at Level 4.
- Access to grade-level literacy standards/objectives – This appears at Levels 3 and 4. This means that the teacher is delivering the grade-level curriculum outlined in the student content standards. This does not mean that students whose skills are far below grade level should be taught material that is beyond their abilities to comprehend or produce. However, the curriculum is not "dumbed down" for these students; addressing critical prerequisite knowledge and skills does not eliminate a focus on more complex knowledge and skills. For example, students who cannot write an essay independently can complete a graphic organizer that helps them organize sentences communicating their ideas into thematic paragraphs within an essay format.

Level differences:

- Between 1 and 2: If there are significant errors in the content being taught, this rubric is scored **at Level 1**. An alternative characteristic of **Level 1** is that aspects of the students' experiential backgrounds, interests, or prior learning are reflected in the plans and have a superficial relationship with the standards/objectives, so the connections aren't very useful in helping students learn the content. In contrast, **at level 2**, this relationship is used in the plan to help move students toward meeting the learning objectives. **At Level 2**, any errors present do not significantly disadvantage students in future learning. In addition, there is at least one general strategy for addressing the needs of students who often have difficulty, e.g., the candidate plans to circulate while students are working and help those who are struggling.
- Between 2 and 3: **At Level 3**, the plans not only draw on students' prior learning, but they also draw on students' experiences or interests to help them meet standards and reach the learning objectives appropriate for their grade level. These candidates *structure* support strategies to help students gain access to the grade-level curriculum. For example, if heterogeneous grouping is planned, there is a process to ensure that students do not just copy the work of others but actively engage in developing their own understanding.
- Between 3 and 4: **At level 4**, the candidate's plans and commentary suggest an understanding of how to meet varied student needs in a classroom. This is not limited to English learners or special needs students, although these students have specific needs that often require differentiation or strategic teaching decisions. The candidate may identify other types of student needs that are being considered during planning, e.g., students who are reluctant to participate in discussions, students who already know the content or who learn it more quickly than other students. There are two approaches to accommodating particular student needs: 1) differentiating instruction, where different instruction is planned to address the needs; and 2) strategic teaching decisions, where instruction is planned that simultaneously addresses multiple needs, perhaps with scaffolding or additional support for students who need it. The candidate need not be meeting *every* student's learning needs, but there should be evidence that there are reasonable strategies for meeting both the needs of students as a class and a variety of distinct needs of individuals or subgroups.

3 Designing assessments

Big ideas and their progression across the rubric:

- Match of standards/objectives, and assessments – This refers to a match between the standards/objectives being assessed and the assessment instruments. It progresses from a mismatch or a limited match between instruction and the learning actually assessed at level 1 to a match at increasingly complex levels from Levels 2 to Levels 3/4. (Remember that candidates were asked to indicate when they were only focusing on part of a standard, so they should be held accountable only for that part, and not the entire standard.)
- Type of student understandings measured by assessments – This idea refers to the complexity of the understandings or skills measured. The assessments move from the

ability to measure at most surface-level student understanding at level 2 to the ability to measure some depth of student understanding at Levels 3 and 4.

- Productive/receptive modalities – Productive modalities are how students communicate their own understanding, while receptive modalities are how students understand communication by others. For example, to assess student understanding of a cause/effect argument, constructing an original speech or essay would assess students' productive understandings while critiquing another's speech or essay would assess their receptive understandings.
- Accommodation of special student needs – This would include any effort to address the needs of a student who otherwise would not be able to demonstrate the relevant knowledge, skills, and understandings on the assessment instrument(s). This could include scaffolding of student work or responses or changes in the method of administration, the format of the assessment, the method of responding to the assessment, or the learning objectives (if reflected in an IEP).

Level differences:

- Between 1 and 2: A significant mismatch between one or more assessments and the content and skills inherent in the learning tasks described in the plans results in **a Level 1 score**. At least one assessment requires knowledge and skills which go *far* beyond those described in the context commentary or taught during the learning segment. The mismatch should be major and should not be confused with reasonable extensions of learning. An alternative way of receiving **a Level 1 score** is that at least one of the assessments does not match the learning objectives identified as being assessed. The candidate may assert that the assessment measures a particular learning objective, e.g., conceptual understanding, but you cannot figure out how. Again, the mismatch should be significant. **At level 2**, the standards/objectives, instruction, and assessments match. However, this match is only clear at a surface-level of understanding.
- Between 2 and 3: **At level 3**, the assessments clearly allow students to display their understanding or skill in some depth relative to the students' developmental level, the short length of the learning segment, and the amount of time students have been working on the particular concept, skill, or understanding as described in the standards/objectives. In addition, both students' ability to communicate their own understandings and skills (productive modalities, e.g., writing, speaking, drawing/graphing) and their understanding of content communicated by others (receptive modalities, e.g., reading, listening, viewing) are assessed.
- Between 3 and 4: **At level 4**, assessments reflect a deliberate design, changes in the assessment instrument or method of administration, or options offered to address the special needs of one or more students who otherwise would be limited in the ability to demonstrate the expected understandings and skills.

Instruction Rubrics

This task provides the only opportunity in the Teaching Event to see the candidate interacting with students. The video clip(s) illustrate how the candidate implements some of the strategies described throughout the Teaching Event and works with students to move them toward meeting standards/objectives. The commentary serves mainly to provide context for understanding the events seen in the clip(s) and for the candidate to point out and explain specific strategies to meet student needs. Information provided in the commentary that does not help the scorer understand and interpret the teaching and learning in the video clip(s) should be given little weight.

4 Engaging students in learning

Big ideas and their progression across the rubric:

- <u>Student opportunities to engage in developing their own understandings</u> – There is a difference between *participation* in learning tasks, i.e., following instructions to complete the activity, vs. intellectual *engagement* with the learning tasks, i.e., actively working with the content throughout the activity so that new learning occurs. The emphasis is on *opportunities* to engage with the content, recognizing that some students may resist engagement. The opportunities vary from limited at Level 1 to opportunities that reflect some explicit tailoring to the students at Level 4.
- <u>Focus of clip(s)</u> – In every content area, the focus of the clip(s) is specified. If the clip(s) submitted do not reflect the appropriate focus, this Guiding Question is scored at Level 1.
- <u>Problematic classroom management that interferes with learning</u> – At Level 1, this is a classroom environment where disruptions or a lack of structure consistently make it very difficult for any student to engage with the content or where the candidate shows so little respect for students and their ideas that the classroom climate is not conducive to student learning.

Level differences:

- <u>Between 1 and 2</u>: **At Level 1**, the students are following instructions and completing the activity, but there is something that limits students' opportunities to develop their own understanding of the required focus of the video clip. This may be due to the content of the questions asked by the teacher or the nature of the activity that the students are asked to do. The teacher may only be interacting with a few of the students without attempting to engage the others. Level 1 also includes candidates whose classrooms are so disruptive or disrespectful of students and their ideas that the environment consistently interferes with student learning, as well as candidates whose video clip(s) do not reflect the required focus. **At Level 2**, the strategies offer students opportunities to *engage* with the content relative to the required focus of the clip(s). Not all students may be actually doing so, but teacher attempts to engage students in the learning task (not just participate) can be identified within the clip(s).
- <u>Between 2 and 3</u>: **At Level 3**, the strategies are structured to engage students intellectually in the learning task(s) and incorporate some attention to students as individuals, i.e., who the students are, their language needs, or other specific learning needs. These strategies may be weakly implemented and/or not explicitly identified by the candidate as intentional.
- <u>Between 3 and 4</u>: In addition, **at Level 4**, candidates need to explicitly identify strategies for intellectual engagement, either in the Instruction commentary or earlier

in the Planning Task. The strategies should be clearly recognizable in the video clip(s).

5 Monitoring student learning during instruction

Big ideas and their progression across the rubric:

- Strategies for monitoring student understanding – Candidates are monitoring student learning at all rubric levels. The candidate's monitoring becomes increasingly sophisticated, with the focus of monitoring deepening until at level 4, they are monitoring, in part, by soliciting explanations of student thinking.
- Candidate responses – This varies from candidate responses that aren't effective at moving student understandings forward at Level 1 to responses at Levels 3 and 4 that build on what students are saying or doing to deepen student understanding of the content and/or thinking processes being taught.
- Significant content inaccuracies – This only appears at Level 1. (If the content is inaccurate, it doesn't matter how it is taught.) Significance is defined in terms of impact on student learning. A significant inaccuracy negatively affects student learning during the instruction viewed on the videotape. Candidates should not be penalized for misspeaking or making minor content errors. The error(s) should be significant enough that the students are going to be disadvantaged when they encounter the same content or apply the same skills again. An example of a significant inaccuracy is defining density or metaphor incorrectly and consistently reflecting this error in the materials and activities seen on the clip.

Level differences:

- Between 1 and 2: Candidates displaying one or more inaccuracies that negatively impact student learning are scored **at Level 1**. Alternatively, candidates who primarily monitor student learning by asking yes/no or other types of simple questions (either orally or through written materials) that don't require much thinking on the part of students (i.e., surface-level questions) are also scored **at Level 1**. In contrast, candidates **at Level 2** are requiring students to think to respond during the activities shown in the video clip(s). This student thinking is grounded in knowledge of facts, skills, conventions, etc., and is not just providing unsupported opinions. Moreover, candidates **at Level 2** respond to students in ways that are "reasonable" attempts to improve student understanding. Reasonable means that candidates are attempting to apply instructional strategies and are making an effort to direct students to some content understanding that requires thinking, not just responding with prior knowledge or just parroting back what has been said previously. The strategies that candidates are using may or may not be working, but their purpose is clearly to get students to think more deeply. These candidates have more to learn about using strategies effectively, but they are making a reasonable effort to get students to think.
- Between 2 and 3: **At level 3**, candidates are using the responses to guide what they do next, in such a way that they are building student understanding. It is evident from candidates' responses that they are evaluating the students' responses and making

decisions accordingly to support students in developing the desired understanding or skills.

- Between 3 and 4: **At level 4**, candidates are making thinking visible so that students understand the reasoning behind at least some responses, modeling thinking processes or helping students understand what is important to notice and talk about in the content area.

Assessment Rubrics

6 Analyzing student work from an assessment

Big ideas and their progression across the rubric:

- Student errors, skills, and understanding − This focus is reflected in both the assessment criteria (including criteria reflected in a rubric) and in the analysis of student performance on the assessment. It ranges from a focus that is off target with respect to the candidate-identified standards/objectives at Level 1 to identifying partial understandings and using patterns in student work as windows into student understanding at Level 4.
- Patterns in student performance − This involves noticing some trends in student performance, either for the class as a whole, for individual students over time, or for subgroups of students. It ranges from unsystematic identification of gross differences between levels of student performance at Level 2 to identifying patterns in both the whole class as well as for either individuals or subgroups in Levels 3 and 4.

Level differences:

- Between 1 and 2: **At Level 1**, either the assessment criteria and/or analysis are not aligned with the standards/objectives identified as being assessed or evidence in the student work samples is not consistent with the conclusions drawn. This may be due to either flaws in the assessment instrument chosen or to flaws in the analysis. An “analysis” that does not address student performance, e.g., a description of instruction leading to the assessment, though not referenced in the rubric, also merits a Level 1 rating. **At Level 2**, the candidate’s analysis is a listing of students’ successes and errors or misunderstandings which are related to the relevant standards/objectives. However, the candidate makes few attempts to use these to understand what the student might have been thinking or doing as they produced their responses. A Level 2 analysis also identifies a few general characteristics of student learning or performances that constitute differing degrees of attainment of the learning objectives.
- Between 2 and 3: **At Level 3**, the analysis uses student errors as an indicator of student understanding. It goes beyond cataloguing successes and errors/misunderstandings on the assessment instrument to describe patterns, either for individuals or for subgroups of students, that shed light on the extent of student understanding or skill. In the case of errors or misunderstandings, the candidates uses patterns to probe for specific sources of misunderstandings, e.g., lack of understanding of a particular concept or procedure, inattention to detail. The pattern

for individuals may be within the work sample or over time, using other sources of evidence and connecting them to the performance in the work sample provided. These patterns are strategically chosen to gain insight into possible intervention points to address student errors or misunderstandings in order for them to make progress relative to the standards/objectives.

- Between 3 and 4: **Level 4** adds partial understandings. The candidate is able to recognize incomplete progress toward the standards/objectives and identify parts that the student has mastered as well as additional parts that the student(s) need to work on. **At Level 4**, the detail and clarity of the analysis indicates a depth of understanding of student performance and more comprehensive consideration of various dimensions of student performance than analyses scored at Level 3.

7 Using assessment to inform teaching

Big ideas and their progression across the rubric:

- Focusing of next steps − This refers to the degree of targeting of the strategies in the proposed next steps. It ranges from strategies that do not focus on student needs identified in the analysis at Level 1 to strategies that strategically target support to specific individuals and groups (the whole class or subgroups) to address specific needs. NOTE: The next steps may address students who may need greater challenge as well as students who are not yet fully meeting the learning goals.
- Basis for next steps − This relates to the breadth of needs that are addressed by the next steps. It ranges from general needs within the class at Level 2 to more specific needs that vary by individual or subgroup at Levels 3 and 4.

Level differences:

- Between 1 and 2: There are three ways to achieve a **Level** 1 rating: 1) the next steps are either vaguely described, e.g., "more support" with no details as to the focus of support or how it would be offered; 2) the next steps are not very closely related to any of the conclusions drawn in the analysis; or 3) the analysis was so flawed that the next steps are not suitable to meet student needs indicated by the student work samples. **At Level 2**, the next steps are based on broadly-defined patterns of performance, and are focused on student misunderstandings, errors, or a need for greater challenge.
- Between 2 and 3: **At Level 3**, the next steps are more targeted to individuals or groups and the needs addressed are more specifically defined. The next steps are based on a deeper level of analysis that distinguishes needs of individuals or subgroups.
- Between 3 and 4: **At Level 4**, the next steps are very targeted, in such a way as to indicate a clear understanding of the key features of content and/or language standards/objectives as well as how to use knowledge about students to help them learn.

Reflection Rubrics

8 Monitoring Student Progress

NOTE: Evidence for this rubric comes primarily from the Daily Reflections.

Big ideas and their progression across the rubric:

- Monitoring student learning − This reflects the attention that candidates give to student learning in their reflections. It ranges from an inconsistent focus on student learning at Level 1 to consideration of the extent to which students are making progress toward meeting standards/objectives at Levels 3 and 4.
- Adjustments in instruction − This reflects the purpose of the adjustments to instruction. It ranges from limited evidence of adjusting instruction to improve student learning at Level 1 to adjustments that are focused on addressing specific learning needs, both for individual students and groups of students, at Level 4. For this rubric, these adjustments should not be only hypothetical, but you should see evidence that they have been implemented (unless the candidate is prevented from making desired adaptations by either the cooperating teacher or district policy; in this case, adjustments will be hypothetical only).

Level differences:

- Between 1 and 2: **At Level 1**, student learning is not consistently monitored. These reflections often make global assertions like "Went well today" without considering if this was true for all students or offering an observation of student performance that suggests what led to that conclusion. Alternatively, candidates may indicate that some students are having difficulty or, conversely, that students are easily learning the material, without considering any implications for the future lessons planned. **At Level 2**, the reflections on student learning resemble a list of what students could or could not successfully do during each lesson. The reflections may also include considerations of time management or problematic student behavior that are independent of the consequences for student learning. However, the modifications of plans is limited either to procedures for implementing activities (e.g., better estimating what can be done during the time period or being more clear about what is needed to complete a learning task) or to going over the same materials in the same way for students who did not understand.
- Between 2 and 3: **At Level 3**, there may also be consideration of the use of instructional time to complete learning tasks, improving directions, or other classroom management issues, but a focus on student progress is also evident. Candidates' reflections are connected across lessons or associated with specific standards/objectives to give a notion of the degree of progress toward meeting the standards/objectives. At least some adjustments to instruction focus on specific learning needs, both for individuals and one or more groups of students (which may include the whole class).
- Between 3 and 4: **At Level 4**, the adjustments are well targeted at features of student learning for the learning segment that are most central in helping students meet the

standards/objectives. These features differ among content areas, but are described in general terms in the rubric.

9 Reflecting on learning

Big ideas and their progression across the rubric:

- Grounding of reflections in research and theory − Candidates are expected to draw upon research and theory to reflect on their teaching practice. This ranges from extremely inappropriate uses of specific research or theory at Level 1 to the appropriate integration of research and theory with knowledge of their students and knowledge of the content area being taught at Level 4.
- Basis for changes in teaching practice − This ranges from proposed changes that reflect assumptions that contradict basic principles of teaching and learning at Level 1 to changes that are appropriate and intentionally targeted at improving the learning of individuals and groups at Level 4.

Level differences:

- Between 1 and 2: One way to score **at Level 1** is to cite a theory or research finding that has nothing to do with the strategy, event, or student performance the candidate is reflecting on (e.g., using Piaget's stage theory as a rationale for group work) or to offer an erroneous interpretation or explanation of a research finding or theory (e.g., an assertion that Bloom's taxonomy suggests that students cannot analyze or evaluate ideas unless they have mastered basic skills in the content area). These errors should be egregious and not subtle. Alternatively, there is little or no evidence that candidates can make appropriate connections between their teaching practice and student learning. At **Level 2**, the reflections are consistent with theory and research, at a general level, but they are not closely connected. The candidate does not seem to be using research and theory to make sense of experience, but more searching for a way to apply familiar research and theory in some fashion. Candidates also identify changes in their teaching practice to solve some problem that they identified. These changes reflect an assumption about how their teaching affected student learning. You may know, based on experience, that either this assumption is not the most likely or the change that they suggest is not likely to work. However, the key idea is that the assumption or the change would seem reasonable to candidates, given their limited experience at this stage of their teaching careers in applying what they have learned.
- Between 2 and 3: **At Level 3**, candidates use principles of theory and research to make sense of what they observed about their students and their learning. This should be explicit (thought not necessarily detailed) in the reflections.
- Between 3 and 4: Compared with a level 3 performance, **at Level 4**, there is a closer connection between the research/theory cited, knowledge of students, and knowledge of content. The changes proposed address the learning of both individuals and groups of students and are tied to the standards/objectives for the learning segment.

Rubrics Based on Evidence Across Teaching Event Tasks

The academic language rubrics differ from the previous rubrics in that they are designed to draw from evidence across all tasks. Note that they focus on academic language both as a medium for learning content and as an independent dimension of content learning. The academic language that is the focus of these rubrics is defined as the language needed to accomplish school tasks, and includes such things as subject-specific vocabulary, grammatical structures that commonly occur in oral and written texts in the content area (e.g., If….then; By …., the author is….), language functions (e.g., predicting, reporting, explaining, convincing), and structures and conventions that are characteristic of types of oral and written texts commonly used in a field (e.g., lab reports, literary discussions, presentations of a problem solution).

While these rubric are critical to instruction of English learners, it also applies to instruction of native speakers of varieties of English, and even speakers fluent in the academic English used in school who will be expanding their command of academic English. However, the candidate should not focus on the language development of fluent speakers of academic English while ignoring English learners and speakers of varieties of English. The overarching idea is the extent to which candidates are making the structure of oral or written texts in the learning tasks transparent to students so that they can better understand and produce that type of text and understand the meaning of subject-specific vocabulary so that they can understand it and use it appropriately. Examples of oral text types include challenging another student's problem solution (for mathematics students) or communicating comparisons for kindergartners. Examples of written text types include cause-effect arguments in history-social science or descriptions of an observation of a scientific phenomenon.

10 Understanding language demands

Big ideas and their progression across the rubric:

- Language demands of learning and assessment tasks – The goal is to move beyond an analysis of vocabulary and grammatical structures in order to focus on (a) the language demands of the learning tasks (e.g. reading written directions, listening to the teacher talk, interacting with students to complete a groups project, completing different kinds of writing tasks, giving oral presentations,) and (b) the features of particular text types (e.g., procedures in a lab report, a "show and tell" presentation, a personal narrative, a justification for a mathematical procedure). Candidates are asked to describe the language demands of learning tasks in the Planning Commentary, but additional information may appear in other places where candidates comment on their students' proficiency with academic language. This ranges from a missing or superficial description of the academic language demands at Level 1 to an analysis that includes vocabulary and text type features at Level 4.
- Students' language strengths and needs – This ranges from general spelling and grammatical errors unrelated to the subject-specific texts used in the learning segment at Level 1, to identification of specific language strengths and needs relative to the

language demands of the learning tasks and texts together with an analysis of what makes the language demanding for individuals or groups at Level 4.

Level differences:

- Between 1 and 2: Level 1 reflects a superficial understanding of language development and language demands. The description of language demands of tasks is either missing or very shallow. The discussion of student language development focuses on errors easily recognizable to non-educators, such as spelling and grammar, but not patterns of spelling and grammatical structures related to the texts used in the learning segment. **At Level 2**, virtually the only language demand that candidates describe that goes beyond grammar and spelling is related to vocabulary, either new vocabulary or different meanings of familiar words.
- Between 2 and 3: **At Level 3**, candidates go beyond identifying vocabulary as the sole language demand to include challenging features of subject-specific text types or more general language demands related to routine tasks in the classroom. (See the accompanying footnote on the rubric for examples.) The discussion of students' language development relative to these demands includes student strengths as well as developmental needs.
- Between 3 and 4: **At Level 4**, candidates also identify and explain features of the text types or other demands that make them challenging for a student or a group.

11 Supporting Academic Language Development

Big ideas and their progression across the rubric:

- Scaffolding or support for academic language – Scaffolds refer to strategies to help students better understand content and understand and/or use academic language. Scaffolds act as a bridge to address gaps between students' current language development and the language demands of the learning and assessment tasks. At Level 1, there is either no scaffolding or support OR language gaps are avoided by oversimplifying language and/or content. At the other end of the range, Level 4, candidates not only provide bridges to help students understand core content but also use strategies to further develop proficiency in academic language and provide rationales for the strategies.

Level differences:

- Between 1 and 2: **At Level 1,** either no language support is provided to address student needs in relationship to the language demands of tasks or the content and/or language is so oversimplified so that little or no development in either content or language takes place. **At Level 2**, candidates identify specific strategies for closing identified language gaps between student levels of development and the demands of learning tasks and assessments. However, while these strategies allow access to content, there is an *absence* of strategies that are specifically targeted at developing language proficiency. Examples of such strategies that provide access to content without developing specific academic language include using pictures in the absence of accompanying language or pairing English learners with a more fluent English speaker who shares the primary language with no provision for the more fluent student doing more than serving as a translator to help the less fluent student through the learning task.

- Between 2 and 3: **At Level 3**, the scaffolds and supports offered to bridge gaps in needs relative to demands not only provide access to content understanding but also target language development. Strategies to help give students understand curriculum content that does not build English proficiency may be present as well to build content understandings. However, at Level 3 there explicit strategies to develop academic language must be present.
- Between 3 and 4: **At Level 4**, there is also an explanation of the instructional strategies and how they address the language gap between what the students can currently do and what the learning task requires them to do.

Primary Sources of Information for Each Rubric
(i.e., candidates were directed to provide information relevant to the rubric)

Context for Learning Task

- Context Commentary – description of students' development – content, language, social skills, and perhaps community resources relevant to the unit

Planning Task

GQ1 (Establishing a Balanced Instructional Focus)

- Lesson Plans (standards, objectives, assessments, instructional strategies and learning tasks, resources)
- Instructional & Assessment Materials (to help you understand the plans in more detail)
- Planning Commentary (especially questions 2 & 4)

GQ2 (Accessible Content)

- Instructional Materials (to help you understand the plans in more detail)
- Planning Commentary (especially questions 2, 3, and 5)

GQ3 (Assessment Design)

- Lesson Plans (standards, objectives, informal and formal assessments)
- Instructional Materials (to help you understand the plans in more detail)
- Planning Commentary (especially questions 4 and 5)

Instruction Task

GQ4 (Engagement in Learning)

- Video clip
- Lesson Plan(s) (to help you understand the clip(s) in the context of the surrounding instruction)
- Instruction Commentary (to help you understand and interpret what you saw in the clip(s), especially questions 1, 2, 3, and 4)

GQ5 (Monitoring Learning

- Video clip
- Instruction Commentary (especially question 5)

Assessment Task

GQ6 (Analysis of Student Work)

- Evaluative Criteria or Rubric
- Assessment Commentary (especially questions 2, 3, 4, 5)
- Student Work Samples (to help you see how candidate actually applied the evaluative criteria and the correspondence of the evidence in student work to conclusions drawn)

GQ7 (Assessment Informing Teaching)
- Assessment Commentary (question 6)

Reflection Task
GQ8 (Monitoring Student Progress)
- Daily Reflections

GQ9 (Reflection on Learning)
- Reflection Commentary (all questions)

Academic Language
GQ10 (Understanding Language Demands)
- Context Commentary, for students' language development (question 2b)
- Planning Commentary, question 3

GQ11 (Supporting Academic Language Development)
- Context Commentary, for students' language development (question 2b)
- Lesson Plans (standards, objectives, assessments, instructional strategies and learning tasks, resources)
- Instruction Commentary, question 4
- Assessment Commentary, questions 5 and 6

Implementing Scorer Training

This section includes a description of the facilities, equipment, and materials needed to conduct training, options for implementing training, potential pitfalls to avoid, and resources available on the PACT website (www.pacttpa.org) to facilitate scoring and scorer training. It is the responsibility of Program Directors to assign responsibilities for implementing training and to ensure that the implementation tasks are carried out in a timely fashion.

Facilities, Equipment, and Materials

The facilities requirements vary with the size of the group of scorers being trained. Facilities should be suitable for the following activities:

- For large groups, viewing PowerPoint slides or overheads
- Small group discussions
- Viewing of video clips
- Sufficient space per person to work simultaneously with multiple documents

Equipment needs include:

- For large groups, projector to display PowerPoint slides or overheads
- Equipment to display video clips, which are provided on CDs (instructions for downloading QuickTime and Windows Media Player are provided in a paper handout with the CDs of the clips).

Small groups can review PowerPoint Handouts of slides and view the videos on a laptop.

<u>Materials</u>
Benchmarks for levels 1, 2, and 3 (Written materials + video clips) - one copy of the written materials per person
Scoring forms with summaries for each of the three benchmarks - one copy per person
Three blank scoring forms (one for each benchmark reviewed) per person
Materials for taking notes - Post-its, note pads, highlighters
A chart labeled "Parking Lot", to be posted on the wall
Post-its for posting onto the Questions/Parking Lot chart
Blank overhead or chart paper for Bias activity
Handouts for each content area scored:

- Agenda
- PowerPoint handout (3 slides per page)
- Teaching Event Handbook
- Relationship of TPEs to Guiding Question Rubrics
- Notes about Bias
- Primary Sources of Information for Each Rubric
- Teaching Practices I Like/Dislike
- Thinking Behind Rubrics

Signs to indicate content area training sites

TIP: Handouts might be organized in pocket folders for each participant, e.g.,

On the left side

Agenda
PowerPoint Handout
Teaching Event Handbook
Relationship of TPEs to GQ Rubrics
Notes about Bias
Teaching Practices I Like/Dislike
Thinking Behind the Rubrics
Primary Sources of Information
for Each Rubric

On the right side

Blank scoring form
Analysis of “2” Benchmark
Blank scoring form
Analysis of “3” Benchmark
Blank scoring form
Analysis of “1” Benchmark

TIP: Copy multi-page materials two-sided or they won’t fit into a folder.

TIP: The text of the three benchmarks is too large to fit into a pocket folder. If the pages of each benchmark are fastened together, e.g., stapled, less time is lost during training due to trying to reorder pages. Fastening the text of all three benchmarks together, e.g., with a large binder clip, facilitates their distribution to individuals.

A two-page checklist of materials that need to be copied for training follows.

Materials To be Duplicated for Scoring Training

Across Content Areas		
Document	**# of copies**	**Made?**
Agenda	1 copy per participant & trainer	
Training Design	1 copy per trainer	
PowerPoint handout (3 slides/page)	1 copy per participant & trainer	
Bias handout	1 copy per participant & trainer	
Bias worksheet	1 copy per participant & trainer	
Thinking Behind Rubrics	1 copy per participant & trainer	
Relationship of TPEs to Guiding Question Rubrics	1 copy per participant & trainer	
Primary Sources of Information for Each Rubric	1 copy per participant & trainer	
For Each Content Area		
Document	**# of copies**	**Made?**
Handbook	1 copy per participant & trainer	
Blank Scoring forms	3 copies per participant + 1 for trainer	
Low benchmark (BM1) text	1 copy per participant & trainer	
Completed scoring form for low benchmark	1 copy per participant & trainer	
Level 2 benchmark (BM2) text	1 copy per participant & trainer	
Completed scoring form for Level 2 benchmark	1 copy per participant & trainer	
EM – high benchmark text	1 copy per participant & trainer	
Completed scoring form for Level 3 benchmark (BM3)	1 copy per participant & trainer	

Options for Implementing Training

Here are two options for conducting the training for the modules:

2 day training
This is the design outlined in the Training Protocol included in this Handbook.

4 modules of 1 ½ - 2 ½ hours
This breaks down the two-day training into four modules, and assumes that the benchmarks are read ahead of time. For the Level 1 and Level 3 benchmarks, it assumes that the benchmarks are read and portions are scored ahead of time. The four modules, labeled within the Training Protocol, are:

1. Overview of Rubrics and Scoring Process, which lasts about 1 ½ hours including an introduction. In the document “Training Protocol for a Two Day Training”, this corresponds to the activities on the first day from 9:00 to 10:30.
2. Understanding Level 2, which lasts about 2 1/2 hours if the benchmark is read ahead of time. In the document “Training Protocol for a Two Day Training”, this corresponds to the activities on the first day from 10:30 to 4:30.
3. Understanding Level 1, which lasts about 2 1/2 hours, if the benchmark is read and scored ahead of time, with the exception of the Instruction task, which requires viewing the video.
4. Understanding Level 3, which lasts about 2 1/2 hours.

Potential Pitfalls

The following are potential pitfalls to avoid in conducting the training:

- Conducting the training long before actual scoring, with no plans for review before actual scoring. If the training is conducted a month or more before scoring, there may be considerable scorer drift. One way to handle this is to select a Teaching Event completed early to score and discuss together, ensuring that all scorers are still calibrated.
- Considering people who only attend part of the training to be trained. Given the schedules of faculty, supervisors, and other potential scorers, it is likely that one or more people will be unable to attend all of the training. The purpose of the training is to familiarize scorers with the scoring process and with the meaning of Levels 1, 2, and 3 for each Guiding Question so they can apply the rubrics to Teaching Events. The training has been carefully designed to accomplish this goal. Therefore, it will be necessary to plan how to acquaint people who do not attend the full training with the material that they missed. Anything that simulates the discussions would be good, because participants reported that that was the most valuable part of the training for them.
- Allowing too many birdwalks during training. The training time is very conservative and does not allow for many related, important discussions which, however, do not bear directly on scoring, e.g., supporting students through completing a Teaching Event, questions about implementation. Writing these questions down and posting them on the Parking Lot with a mechanism to address them at a later time (either briefly, within the training, or in a Q and A document) is necessary to keep within the allotted time.
- Managing debates on benchmark scores. Sometimes the performance has characteristics that match two adjacent level descriptors, i.e., it is borderline with respect to a specific Guiding Question. In this case, professional judgment is used to select a single score point based on a preponderance of evidence. Professional judgments may differ when the performance is truly near the border. Agreement should not be forced in those cases; the borderline nature of the performance and the difficulty in reaching agreement in those cases should be acknowledged. Teaching Events that are scored near the passing point will be read multiple times so that passing does not depend on one scorer's judgment.
- Being too flexible on accepting different professional judgments. Arguments for a particular score point should be firmly grounded in evidence from the Teaching Event, with specific references to features of artifacts or quotes from the commentaries. Key features of rubric level descriptors (which are bolded) should be addressed in arguing for one score vs. another. Arguments which rely on evidence that maps to characteristics that are not reflected in the rubric level descriptors should be recorded as a suggested change in the rubrics and posted on the Parking Lot chart. Those suggestions should then be shared with the PACT staff at Stanford University for consideration in the next round of rubric revision.
- Being too inflexible on revising benchmark scores. Sometimes one or more scorers offer additional evidence or a reinterpretation of the evidence already noted that clearly matches a score other than the assigned score. Their argument

may be more compelling that the justification of the original score. In that case, the trainer should acknowledge the superior argument and communicate both the evidence and the rationale for the new score to the PACT staff at Stanford. Trainers are not being asked to defend the indefensible.

Resources Available on PACT Website

www.pacttpa.org

Under TE Materials tab

- Teaching Event Handbooks for all ten subject areas
- Making Good Choices, a resource guide to help students understand the Teaching Event
- PACT Videotaping Policies outlining the use of the Teaching Events
- PACT Videotaping Tips
- TPEs
- PE Model Content Standards

Under Rubrics tab

- Rubrics for all ten subject areas

Under Required Forms tab

- Teaching Event Authenticity Form
- Candidate Survey (forthcoming)

Under Scoring tab, and Scoring Materials (to be updated soon)

- Handbook for Implementing Scoring Training
- Thinking Behind the Rubrics
- GQ-TPE Map

Under Scoring tab, and Required Forms (to be updated soon)

- All scoring forms