Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synopsis</td>
<td>15.i</td>
</tr>
<tr>
<td>Introduction</td>
<td>15.1</td>
</tr>
<tr>
<td>An early decision about state and district measures of learning</td>
<td>15.1</td>
</tr>
<tr>
<td>A late decision to expand observation-based measures of student</td>
<td>15.1</td>
</tr>
<tr>
<td>engagement in learning</td>
<td></td>
</tr>
<tr>
<td>The TEP-2 teacher effectiveness portfolio</td>
<td>15.2</td>
</tr>
<tr>
<td>Measures pertaining to student learning obtained through the</td>
<td>15.2</td>
</tr>
<tr>
<td>TEP portfolio</td>
<td></td>
</tr>
<tr>
<td>Scoring TEP portfolios for a teacher’s impact on learning</td>
<td>15.3</td>
</tr>
<tr>
<td>Trustworthiness of TEP portfolio-based measures pertaining to learning</td>
<td>15.4</td>
</tr>
<tr>
<td>Measures of teacher impact on learning obtained through classroom</td>
<td>15.6</td>
</tr>
<tr>
<td>observation</td>
<td></td>
</tr>
<tr>
<td>Trustworthiness of the observation-based measures pertaining to learning</td>
<td>15.7</td>
</tr>
<tr>
<td>Summary</td>
<td>15.8</td>
</tr>
<tr>
<td>ATTACHMENTS</td>
<td>15.9</td>
</tr>
</tbody>
</table>

Prepared by

Del Schalock
Andrew McConney
Mark Schalock

Teaching Research Division
Western Oregon University
Issues Addressed

1. What assessments were made of the impact of teachers taking part in the research on the learning of their students?
2. How, when, and by whom were these assessments made?
3. Is the information coming from these assessments of high quality, trustworthy (reliable, valid), and meaningful?

Content and Organization

Research Brief 15 opens with an overview of the multiple approaches considered for assessing the impact of participating teachers on learning, and why two widely advocated approaches were not used in the TEP-2 research project. These were
- looking to student performance on state or district administered examinations; and
- looking to school or district records of student progress toward benchmarked standards for learning.

The brief then goes on to describe the two other methodologies used in the research to assess teacher impact on learning, and the specific measures derived from each methodology. The two methodologies were
- An extended form of teacher work sampling referred to as a Teacher Effectiveness Portfolio (TEP); and
- Classroom observation, including a feature added to the observation protocol used in the study (see Research Brief 5 distributed to Design Team members in March) that captures the kind and level of intellectual work in which students are engaged during the course of the instructional periods observed.

Both methodologies are described in sufficient detail for readers to understand their nature and use, and the specific indicators of a teacher’s impact on learning coming from them. Evidence also is provided as to the confidence that can be placed in the measures derived from these two methodologies.

Conclusions Reached Around Portfolio Use

✓ The preparation of the portfolio was viewed as a heavy burden by both the 1st and 2nd year teachers taking part in the research (3rd year teachers were relieved of the burden). Thirty five of the 86 1st year teachers who signed on to be a part of the study asked to be released from the research by the completion of their second year of teaching, with 14 of these 35 indicating that participation was simply too much work on top of their teaching responsibilities. The only burden of consequence from participating in the research was portfolio preparation (see Research Brief 1 for other factors contributing to participant attrition).

✓ One factor contributing to low profile completion was the lack of consequences for deciding not to. No “high stakes” consequences were attached to either profile completion or project participation, and the only incentive for either -- beyond personal interest or a sense of professional responsibility -- was a $500 yearly stipend.

✓ We chose to provide minimum structure to the focus and content of the portfolio that was to be prepared (see ATTACHMENT A in this brief) because we wanted to see what early career teachers prepared under differing program emphases would produce under their own initiative. We did not tell them, for example, what criteria would be used in evaluating (scoring) the portfolios, or what measures would be derived from them. By pursuing this strategy “we got what we got”, and this is probably not a fair estimate of what participating teachers were capable of providing. Even so, the over-all quality of the vast majority of portfolios prepared was remarkable. Over and over we were impressed by the cases built and the evidence provided by these early career teachers that “non-trivial” learning was occurring in their classrooms.
Conclusions Reached Around Observing Student Intellectual Engagement

✓ Looking to the kind and level of intellectual work a teacher has students pursue is an appropriate indicator of a teacher’s impact on learning.

✓ Trained observers are able to differentiate between kind/levels of intellectual work in which students are engaged during the course of an instructional period, and estimate the proportion of time spent doing so.

✓ Student “intellectual engagement profiles” can be developed for individual teachers from either of these sets of observational data, one based on the frequency with which a particular level of intellectual engagement is pursued over a series of classroom observations, and one based in the proportion of class time spent in such engagement.

Potential Implications for CONTINUING Licensure

1. The teacher effectiveness portfolio (TEP) design used in the research project has features that hold promise for portfolio design in the CTL process, but since that process leads to inordinately high-stakes decisions clarity and specificity as to scoring criteria (scoring rubrics, performance standards) should be a part of the directions provided.

2. The rubrics and performance standards used in scoring the TEP portfolios for purposes of the research project (see ATTACHMENT B in this brief) could be used meaningfully in evaluating portfolio-based evidence in the CTL process.

3. The observation-based measures of teacher impact on learning described in this brief could also be used meaningfully and defensibly in the CTL process.

Potential Implications for INITIAL Licensure

1. The portfolio-based measures of teacher impact on learning alluded to above could be applied to teacher work sample evaluation for purposes of INITIAL licensure.

2. The observation-based measures of teacher impact on learning alluded to above could be applied to teacher work sample evaluation for purposes of INITIAL licensure.
Introduction

The TEP-2 research project was designed to trace as fully as possible the effects of selected features of preparation programs on the thinking, feeling, performance and effectiveness of early career teachers, and how these vary with time and context. With this comprehensive focus a crucial dimension of the research has been the impact of early career teachers on the learning of their students.

As initially planned the project was to focus on three separate sources of evidence pertaining to student learning:

- Performance on state and/or district administered examinations;
- Teacher documented progress in the learning of their students through an extension of Oregon’s approach to teacher work sampling (the TEP Portfolio);
- A description of the intellectual challenge involved in classroom work pursued by students that occurred during the classroom observations made as part of the research.

Early Decisions About State and District Measures of Learning

As we proceeded with selecting participants for the project, it became clear that Oregon’s design for standards-based schools, with standards for learning benchmarked at designated grade levels and the state’s assessment system tailored accordingly, would not permit all of the elementary teachers taking part in the research to have access to state collected information on the learning progress of their students toward standards for learning. With research participants scattered across grades K through 6 it was not appropriate to use state assessment information as evidence of the impact of participating teachers on the learning of their students.

As the project progressed it also became clear that district supplementary assessment systems varied widely, from none to a Fall-Spring testing schedule each year. It became equally clear that district maintained records of student progress toward standards for learning varied greatly in both content and utility, and that teacher collected samples of student work were dealt with by schools in a myriad of ways. Additionally we found that few districts currently disaggregate their data on learning in such a way that permits them to track teacher effects on individual students across time.

This lack of comparability across the elementary grades in state and district managed information on student progress toward benchmarked standards for learning, and thus across the K-6 teachers taking part in the research, meant that we had to set aside our efforts to collect such information as a central focus of the project. Instead we have relied on the two other sources of evidence pertaining to teacher impact on learning that were part of our initial planning: classroom observations and teacher effectiveness portfolios.

A Late Decision to Expand Observation-Based Measures of Learning

As the classroom observation protocol evolved, and the richness of information provided through it became evident (see Research Briefs 5, 6, 7 and 10), we opted to treat three dimensions of this information -- instead of only one as initially planned -- as indicators of a teacher’s impact on learning. In addition to the kind/level of intellectual engagement pursued by students in a classroom we also decided to look at

- the engagement of students with content, and
- the classroom as an environment for learning.

The remainder of this brief is given to a description of these multiple sources of evidence pertaining to student learning, and the confidence that can be placed in their use.

The TEP-2 Teacher Effectiveness Portfolio

The teacher effectiveness portfolio is a vehicle by which teachers can convey to others a “snapshot” of what they do in a classroom, what they hope to accomplish with students, and what they actually accomplish with those they teach. It is an extended version of Oregon’s design for a teacher work sample, which has been recognized as a practical, defensible, and meaningful way for prospective teachers to demonstrate their ability to connect their teaching to student progress in learning.
The TEP methodology is essentially a teacher work sample that has been refined and extended for use by full-time teachers. It retains most of the elements contained in work samples required of student teachers, but calls for two related units of study rather than one. The TEP methodology also differs from student teacher work samples in that it does not carry the specificity of most work sample prompts used at that level of preparation, and it does not have “high stakes” implications for licensure related decisions. Guidelines for TEP preparation, for example, do not include reference to the dimensions on which they will be scored, and the only incentive for their preparation was the participating teacher’s commitment to the project and the small stipend that accompanied participation. The guidelines provided teachers participating in the project for TEP preparation are appended as ATTACHMENT A.

Participating teachers were given these guidelines early in their 1st and 2nd years of teaching, and asked to submit their completed portfolios before the close of school each spring. Each portfolio was to be presented in a 1½ inch by 3-inch ring binder with side tabs provided by the project. Beyond the binder and guidelines, however, and an opportunity to clarify questions about them in an initial day of orientation to the project and through subsequent telephone or email inquiries with project staff, no assistance was provided in portfolio preparation.

**Measures Pertaining to Student Learning Obtained Through the TEP Portfolio**

The contents of each portfolio were rated on ten dimensions, nine of which pertain to teaching and learning and one to the overall quality of portfolio preparation. The scoring guides used in making these ratings, and the directions provided raters, are appended as ATTACHMENT B.

For purposes of the research project five of the nine substantive ratings made are treated as either direct or indirect measures of a teacher’s impact on the learning of students. Three of the five ratings provide evidence pertaining to teacher documentation of student progress in learning. These are:

- the sufficiency and quality of assessment(s) used to determine student progress in learning (Dimension I5 in ATTACHMENT B);
- the documented progress made by students in their learning (Dimension I6); and
- a teacher’s reflection on his or her practice in light of learning gains achieved (Dimension I7).

Each of these dimensions is made up of constituent parts that are considered separately in a related scoring rubric. Each part receives a score of 1 to 4 (not met, partially met, etc), with these scores being summed as a total score for the dimension being rated. For example, since there are seven parts to the assessment dimension, the scale score possible for assessment is 28 (7 x 4). The scale score possible for the outcomes dimension is 20 (5 x 4), and the reflection dimension 12 (3 x 4).

Two of the five portfolio-based measures pertaining to student learning, however, take a different form than the three just described. These are

- Dimension II, which involves an over-all portfolio rating pertaining to the fostering of non-trivial learning; and
- Dimension III, which involves an over-all portfolio rating pertaining to attention given to Oregon’s benchmarked standards for learning.

Both of these measures require summative ratings of portfolio evidence pertaining to student learning, and receive scores of 1 to 4. These two approaches to scale score derivation are illustrated in ATTACHMENT C, and discussed in greater detail in subsequent pages.

For purposes of the research project we look to all five interrelated measures when considering teacher impact on student learning.

**Scoring TEP Portfolios for Teacher Impact on Learning**

Portfolios were scored independently by one of seven Teaching Research faculty members serving as staff to the research project. As each portfolio
was read all ten of the dimensions described in ATTACHMENT B were scored. Judgments pertaining to student learning were made in light of all evidence presented within a portfolio.

Several half-day periods were devoted to pilot testing, refining, and determining preliminary rater agreement (inter-rater reliability) scores with the portfolio rating system prior to raters scoring the portfolios independently. These preliminary training and scoring guide-refining sessions were carried out on selected “pilot” portfolios until there was general agreement across raters on all rubrics, scoring procedures, and decision rules.

Once this level of refinement and agreement was reached the final scoring of all 1st year teacher portfolios, and all second year portfolios for the first cohort of participating teachers, was completed on two consecutive days. The portfolios used for training and refinement purposes were subsequently re-scored by four of the six staff members participating in the all-day scoring session. As a first step in scoring, seven portfolios were selected and read by three independent raters to fine tune agreements and provide inter-rater reliability data.

Scoring for teacher documentation of student progress in learning. It will be recalled that these three scoring rubrics involve constituent parts, and that a judgment needs to be made about each part being met or not met. The scoring rubrics containing these parts are described in ATTACHMENT B.

Judgments about each part involve a 4-point scale, or a designation that related content has not been provided. The scale descriptors for constituent parts are

1. Does not meet standard;
2. Partially meets standard;
3. Meets standard;
4. Exceeds standard;
99. Related content not provided.

The format used in recording these judgments is appended as ATTACHMENT C.

After making the part-by-part judgments involved in a particular measure, which are then summed to a scale score, an over-all rating of 1 to 4 was provided for the dimension as a whole on the basis of the profile of judgments made about its constituent parts. In arriving at this over-all rating a preponderance of evidence rule was in effect. For example, the assessment dimension contained seven parts. If a teacher’s portfolio addressed all 7 parts, and all parts were judged to have met the standard expected, the summative rating for the dimension as a whole would be 3. If 4, 5 or 6 of the 7 parts were addressed, and all were judged to have met the standard expected, the summative rating for the dimension as a whole would be 2.

This “double scoring” procedure provided two different, through closely related measures of the portfolio-based dimensions of ASSESSMENT, OUTCOMES, and REFLECTION. One involves an additive score for the parts documented (in the case of assessment cited above, this can range from 4 to 28 since seven parts make up the dimension and each part can carry a score of 1 to 4). The other involves an over-all rating for the dimension based on the profile of parts recorded by the portfolio rater, as illustrated in ATTACHMENT C. Both measures are used in the analyses reported in subsequent briefs.

Scoring for over-all evidence pertaining to student learning. As explained previously the portfolio-based measures pertaining to THE FOSTERING OF NON-TRIVIAL LEARNING, and ATTENTION TO OREGON’S BENCHMARKED STANDARDS FOR LEARNING, differ markedly from the three just described. Neither of these measures carries the designation of specific parts. They depend instead on an interpretation of all evidence provided within a portfolio that pertains to learning. As such, they involve broadly summative judgments, with ratings ranging from 1 to 4. The performance standards against which these ratings were made also will be found in ATTACHMENT B.

Both summative ratings are provided after all other dimensions of a portfolio have been scored.
Trustworthiness of TEP Portfolio Based Measures Pertaining to Learning

The confidence that can be placed in any portfolio-based measure of a teacher’s performance rests on many factors. Five play a particularly important role:

- The clarity with which the performance of interest has been defined;
- the means by which performance is captured and presented in the portfolio;
- the representativeness of the performance reported;
- the quality and meaningfulness of the scoring rubrics used to evaluate the evidence presented; and
- the clarity and defensibility of the rules and procedures followed in making the evaluation.

Confidence in portfolio-based measures, however, also depends on empirical evidence that two or more independent judges are able to agree on the evaluative decision that is to be made on the basis of the information provided. Inter-rater reliability data is viewed as obligatory for any rating system leading to evaluative decisions based on the array of disconnected information typically found in portfolios designed to document multiple dimensions of performance.

As stated previously, an inter-rater reliability study was conducted for the seven TR faculty members who scored the TEP portfolios, and the results of this study are presented in the pages that follow. Findings are presented under two headings: Rater agreement across portfolios, and extent and severity of scoring bias.

Rater agreement across portfolios. Agreements among and across sets of 3 independent raters scoring 7 portfolios provide the basis for the inter-rater reliability reported on the next page. These inter-rater-reliability studies were conducted after several half-day sessions were devoted to refining scoring rubrics and decision rules, and occurred immediately prior to final portfolio scoring. Discussions on individual ratings took place to further build common understanding. No additional training of raters occurred after these studies were completed.

The seven portfolios selected for the inter-rater agreement studies included two that on first reading appeared to be outstanding, two that appeared to be weak, and three that appeared to fall somewhere in between. We needed to know that our scoring was reliable across the spectrum of portfolios received, though we suspected we would be less reliable on portfolios falling in the middle range of thoroughness and quality.

The summative ratings provided by each rater within the three member rating teams for each of the five portfolio-based measures pertaining to student learning are presented in Table 15.1. The data provided consists of 105 two-rater comparisons. Inter-rater agreements were not calculated for each of the constituent parts of the ASSESSMENT, OUTCOMES, and REFLECTION measures, though the information needed to do so is available for inspection from the portfolio scoring records.

We interpret the data presented in Table 15.1 as providing an acceptable level of evidence supporting the reliability of our seven portfolio raters on measures pertaining to student learning. Three patterns in the data support this interpretation:

- All but 3 of the 105 paired ratings were in perfect agreement (54), or varied by only 1 scale position (51);
- Raters consistently differentiated between teachers as to levels of performance on the five learning related measures (the sum of ratings appearing on the right hand side of the table), though only one of the two portfolios initially thought to be weak (P124) turned out to be so;
- While there is a tendency on the part of some raters to be more or less generous in their ratings than others they are not always so, and not so much so as to invalidate the scores of any particular rater. These differences are discussed more fully in subsequent paragraphs.

As anticipated perfect agreement between raters was greater when scoring strong or weak portfolios (28 perfect agreements for portfolios 124, 129 and 133) than when scoring portfolios falling between these two extremes (26 perfect agreements for portfolios 125, 127, 149 and 281).
Table 15.1 Inter-rater agreements on the summative ratings provided for each portfolio-based measure pertaining to student learning

<table>
<thead>
<tr>
<th>Portfolio and Raters</th>
<th>Teacher Documentation Measures</th>
<th>Portfolio Summary Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASSESSMENT</td>
<td>OUTCOMES</td>
</tr>
<tr>
<td>P 124 Raters A-B</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A-C</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B-C</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P 125 Raters D-E</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>D-F</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>E-F</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P 127 Raters D-C</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D-G</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>C-G</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P 129 Raters B-G</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>B-F</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>G-F</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>P 133 Raters E-C</td>
<td>*3</td>
<td>1</td>
</tr>
<tr>
<td>E-D</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>C-D</td>
<td>*1</td>
<td>3</td>
</tr>
<tr>
<td>P 149 Raters E-B</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>E-F</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B-F</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>P 281 Raters E-G</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>E-F</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>G-F</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

• One of three inter-rater agreements varying by 2 scale positions

Table 15.2. Direction and extent of scoring bias on the part of individual raters

<table>
<thead>
<tr>
<th>Rater and # of ratings</th>
<th>Direction In Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>A (of 10)</td>
<td>4</td>
</tr>
<tr>
<td>B (of 30)</td>
<td>10</td>
</tr>
<tr>
<td>C (of 29)</td>
<td>10</td>
</tr>
<tr>
<td>D (of 30)</td>
<td>11</td>
</tr>
<tr>
<td>E (of 40)</td>
<td>10</td>
</tr>
<tr>
<td>F (of 40)</td>
<td>5</td>
</tr>
<tr>
<td>G (of 30)</td>
<td>2</td>
</tr>
</tbody>
</table>

Extent and severity of scoring bias. When the agreements and disagreements of individual raters are calculated across portfolios scored, and put in tabular form, it is clear that three of the seven raters (B, C, D) were less generous in their ratings than the other four. Data portraying the scoring patterns of individual raters are presented in Table 15.2.

From the data presented in Table 15.2 it can be seen that a larger proportion of the scores provided by raters B, C, and D were lower than those provided by their co-raters scoring the portfolios while raters A, E, F and G provided a larger proportion of higher scores than those provided by their co-raters.
Both of these directional trends can be interpreted as “rater bias”, and could be cause for concern if they were more pronounced. Since they are not severe, however, and since all raters provided both high and low ratings -- and all raters had perfect agreement with their co-raters on at least half of all ratings made on the 5 measures of interest -- the extent and direction of scoring bias does not seem to be sufficiently great as to be a cause for major concern.

**Measures of Teacher Impact on Learning Obtained Through Classroom Observation**

As indicated previously we believe that three aspects of the information obtained through the TEP-II classroom observation protocol can serve as important indicators of a teacher’s impact on learning. These are:
- the kind/level of intellectual engagement students are asked to pursue in their classroom work;
- the engagement of students with content; and
- the classroom as an environment for standards-based learning.

The rationale for and procedures used in obtaining these three lines of evidence are discussed separately in the paragraphs which follow.

**Levels of student intellectual engagement.**

Looking at the performance of students on state, district and teacher administered examinations represents one approach to addressing the impact of teachers on learning. A complementary, and in many ways more basic approach, is to look at the kind and level of intellectual work students are asked to pursue during the course of an instructional period. Engaging in recognition, recall or memorization activities during most of an instructional period is likely to have a far different impact on long term learning than spending that time in activities focusing on understanding, explanation, or interpretation.

The classroom observations made of teachers participating in the research provided a means of collecting this kind of information. Half-day observations were scheduled twice a year as part of the over-all research design (see Research Briefs 1 and 5 reported in March), and provisions were made to include in the observation protocol a description of the intellectual challenge involved in the classroom work students were asked to pursue during each instructional period observed.

While we were clear about wanting to collect this kind of information, we were not as clear about how best to collect it. As a consequence our methodology for doing so evolved throughout the project. In the first year of data collection we simply asked our observers to describe on the cover page of the observation protocol the instructional activities in which students were engaged during the observation period. Our intent was to classify these descriptive statements through established content analysis procedures to determine the intellectual challenge involved.

In the second year of the study we decided to give added focus and clarity to this aspect of the data collection process by having observers record not only the learning activities in which students were engaged, but also to categorize and note the level of intellectual demand involved. To bring consistency to this categorization process six kinds (levels?) of intellectual engagement were listed in the space provided on the observation protocol for classroom activity description, and observers circled the categories of intellectual demand that accompanied the various learning activities recorded. These category listings are included in the protocol cover page appearing as ATTACHMENT D.

In preparing for the third year of data collection in the project still another refinement was made to this aspect of the observation system used. In addition to classifying and noting the kind of intellectual activity in which students were engaged observers also were asked to estimate the proportion of the instructional period spent in each activity noted. This additional task for observers did not add appreciably to their load, and it provided a much clearer picture of the intellectual demands of a classroom than simply noting the kinds of intellectual activity occurring.

Since data reported in Briefs 17 through 20 are limited to data collected during the first two years of the study (data on 3rd year teachers are still being collected), and we have not as yet conducted the content analyses of the learning activities described in years 1 and 2, we will report in these briefs only
the circled categorical notations of intellectual activity students pursued during year 2 observations. These data represent the number of times each level of intellectual activity was noted in the classroom of a particular teacher during the four instructional periods the teacher was observed in his or her 2nd year of teaching.

Each participating teacher was observed during two instructional periods in the fall and two in the spring. For each observation each category of intellectual engagement to be coded 1 (observed) or 0 (not observed). This means that the range of scores an individual teacher can receive for any particular category of intellectual engagement will vary from 0 to 4, and that a teacher’s profile of intellectual engagement on the part of students will vary from 0 to 4 across all six types of intellectual engagement that could be circled by an observer.

The engagement of students with content. This dimension of the observation protocol focuses on behaviors and practices of a teacher that center on “aligning and varying content to support students in their learning” (see ATTACHMENT E). Observers indicate whether any of six designated aspects of a teacher’s performance in this regard were observed during the course of an instructional period:
- effectively and accurately communicates content;
- varies content to accommodate differences in learners;
- connects content to real-life situations;
- connects content across disciplines;
- engages students in multiple-level and/or higher order thinking tasks involving content;
- identifies and addresses misconceptions about content

A teacher who attends to all or most of these aspects of content related instruction is likely to have a far different impact on learning than a teacher who fails to do so.

The classroom as an environment for standards-based learning. This dimension of the observation protocol takes the same form as the engagement of students with content (see ATTACHMENT E), but the focus of the items within it is on “creating a classroom environment that supports students in their learning.” In this section of the protocol observers indicate whether any of nine aspects of a teacher’s performance in this regard were observed during an instructional period:
- maintains a positive, learning focused classroom environment;
- monitors and manages individual and group behavior to maximize learning for all students;
- manages time and resources to maximize learning for all students;
- quickly and effectively resolves conflicts between students if they arise;
- creatively interweaves content, activities, and discourse to engage students in their learning;
- students help other students in their learning, and help their teacher as needed or requested;
- students engage in self-directed learning;
- students ask clarifying questions;
- physical features of the classroom provide an inviting and supportive context for learning.

As in the case of engaging students with content a teacher who attends to all or most of these aspects of classroom management is likely to have a far different impact on learning than a teacher who fails to do so.

Trustworthiness of the Observation-Based Measures of Impact

To date we have no formal evidence as to the trustworthiness of either the observers’ notations of the intellectual activity pursued by students during the instructional periods observed, or the instructional behaviors recorded around engagement with content and classroom management. The progressive refinements made to the observation protocol each year around these various sets of data were added just prior to the time fall observations were to begin, and no time remained for related observer reliability studies to be carried out. Observers felt comfortable in making these categorizations however (they suggested the category labels and dimensions of behavior ultimately agreed upon), and, after extended discussion leading to the labels and behavioral descriptors selected, expressed confidence in their shared accuracy of use. All sixteen observers employed in the research were experienced teachers, and had been selected by OEA and TSPC staff for the role of classroom observers in the research project because of their status as master teachers and their recognized excellence as
supervisors of beginning teachers. Observers also were rotated each year across teachers. Even with this design enhancement, however, and the level of experience and expertise on the part of our observers, the fact remains that no systematically collected evidence exists on the confidence that can be placed in the categorical (frequency count) data reported from these observation-based measures of teacher impact on learning. Readers need to view the data accordingly.

**Summary**

Our search for defensible indicators of a teacher’s contribution to student learning led us down paths not fully anticipated when designing the TEP-2 project. Our commitment to *multiple sources and multiple indicators* for doing so, however, proved to be sound. The sources and indicators finally adopted, as described in the previous pages, provided a reasonably cohesive and wide-ranging set of lenses for looking at the many dimensions of this complex issue, and as such provides a framework -- as well as illustrative measures -- for CTL program faculties and candidates to draw upon when addressing related issues in the CTL licensing process.
ATTACHMENT A

GUIDELINES FOR PREPARING A TEACHER EFFECTIVENESS PORTFOLIO

Purpose of the Portfolio

There are three purposes for asking you to develop this portfolio:

1. We want to document and make observable one snapshot of what you do as a teacher - how you plan, how you implement your plans, how you assess students, etc.

2. We want to capture your thinking around these things.

3. We want to capture the learning of your students on outcomes reflecting what you actually teach in the classroom rather than on some standardized assessment measure.

What is Important, What is Not

We are interested in the contents of your portfolio. Formatting is secondary.

We do not expect you, nor want you, to spend a lot of time making it look a certain way. Any graphics, tables, charts, etc. do not need to be computer generated, as long as they are legible.

If you have said what you want to say in one page, do not feel that you have to fill up additional pages just to meet the suggested/possible length of narratives. Longer is not necessarily better.

General Contents of the Portfolio

Here are some guidelines and considerations that will help you in assembling the materials for your portfolio. As we mentioned earlier, many of the materials that you will ultimately select for inclusion in your portfolio are those that already exist as part of your, and your students’, work in the classroom. We will refer to these already existing materials as “artifacts”. Artifacts will include, but not be limited to, examples of district curriculum guides or materials used in your work, lesson plans you create for your classes, examples of assessments you use (both formal and informal), examples of student work, examples of communication between you and your students’ parents, notes to yourself, and so on. All of these and other materials can serve as information that, when placed together and organized and reviewed, provides a look at what a teacher does when teaching.

We also ask for your voice in creating some materials for inclusion in the portfolio. We will ask you to write about what you did, observed, planned, felt, and thought about while doing this teaching. Most of what we will ask you to write will be specific to particular parts of the portfolio and each piece of writing should be about 3 pages or less.

General Structure of the Portfolio

Each Teacher Effectiveness Portfolio should be built around a sequence of two or more instruction/assessment units, spanning approximately 4 consecutive weeks of classroom instruction. As stated in our letter we ask you to compile material that represents a substantial and cohesive body of teaching and learning. You may select any content (although the units should be constructed around goals derived from the Oregon standards for benchmark grades 3 or 5). The units could encompass a traditional or integrated unit(s) of instruction, a major
project leading to a student product and/or presentation, or some other piece of instruction.

It is important that the body of work you develop and document be one that includes some type of pre-instruction assessment that shows what your students already know and can do prior to your teaching and then culminates in some post-instruction assessment of your students’ progress. These assessments might include the formal assessment of a product or a presentation using a scoring guide, or it might include some other form of assessment, such as a written test.

**Specific Contents of the Portfolio - Components/Structure**

**Setting/Context**

We already collect demographic information from you on the form that is part of your packet of attitude surveys. What we want to know here is whether there is anything about your classroom, school or community that you think has a strong impact on what you do as a teacher.

**Your Voice** - A brief written narrative describing any particularly important aspects of the classroom, school, or community context in terms of what you do as a teacher.

**Planning Aspects--General**

What we are looking for here is a brief description of the unit(s) and the rationale behind why this unit was developed.

**Your Voice** - A not more than one page written description of why you developed the unit(s), their relevance to each other if appropriate, and their relevance to the district and state standards.

**Artifacts** - Copies of district’s curriculum relevant to teaching units, relevant state content standards, etc.

**Planning Aspects--Specific**

Here we are looking for more specific examples of your actual planning. The kinds of activities you include, the things you take into account in putting together a lesson or series of lessons.

**Your Voice** - A not more than two page written discussion of the rationale for the planning activities and the lessons developed and any specific things you considered when putting these plans together (for example, students’ levels of skills/ knowledge PRIOR to instruction, planning for transitions between activities, group versus individual work, etc.).

**Artifacts** - Copies of the lesson plans developed, the pre-assessment data obtained, notes to yourself, handouts constructed, etc.

**Assessment(s)**

Here we are interested in knowing how you go about knowing what your students understand and how you use that information.

**Your Voice** - A short - probably not more than two pages - written description of the development and/or use of the assessments procedures or tools you provided, what these told you about your students’ learning, and how you used the information you obtained.
Artifacts - Copies of pre/post assessments developed or used, quizzes or other informal assessments employed, scoring guides you provided.

Implementation

Here we are interested in how your lessons went. Did they go as planned? Did you have to change in mid-stream? Tell us what happened and why.

Your Voice - A short paragraph around each lesson describing how the activities were implemented, how students received them, and any modifications you made and why.

Artifacts - Samples of lesson plans and/or lessons implemented, examples of student work, notes on adjustments or modifications you made during the instruction, etc.

Outcomes

Here we are interested in how the unit went from the perspective of what your students learned.

Your Voice - A not more than three page written discussion of student progress and learning during the units. Include a re-examination of the teaching/learning context and its effects on the assessment results and instructional outcomes, an analysis of the assessment data for each student or groups of students within the class, a discussion, if appropriate, of the use you made of informal assessment procedures, and a discussion of your plans for re-teaching these units.

Artifacts - Graphs of student performance from pre-to-post assessments, copies of grading roster for the unit, examples of student work on assessment activities or other indicators of student progress.

Reflection

Your Voice - A not more than three page written reflection on the overall teaching experience you have documented, the thinking, planning, and adjusting you did while developing and implementing the components of the instructional unit(s). When (if) you teach this unit again, how will you do it next time? What did you learn about yourself as a teacher and your students as learners from this sample of your teaching?
<table>
<thead>
<tr>
<th>Portfolio Contents</th>
<th>Artifacts</th>
<th>Teacher Produced for Portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting/Context</td>
<td>A brief written narrative describing any particularly important aspects of the classroom, school, or community context in terms of what you do as a teacher.</td>
<td></td>
</tr>
<tr>
<td>Planning Aspects--General</td>
<td>Copies of district’s curriculum relevant to teaching units, relevant state content standards, etc.</td>
<td>A not more than one page written description of why you developed the unit(s), their relevance to each other if appropriate, and their relevance to the district and state standards.</td>
</tr>
<tr>
<td>Planning Aspects--Specific</td>
<td>Copies of the lesson plans developed, the pre-assessment data obtained, notes to yourself, hand-outs constructed, etc.</td>
<td>A not more than two page written discussion of the rationale for the planning activities and the lessons developed and any specific things you considered when putting these plans together (for example, students’ levels of skills/ knowledge PRIOR to instruction, planning for transitions between activities, group versus individual work, etc.).</td>
</tr>
<tr>
<td>Assessment(s)</td>
<td>Copies of the pre/post assessments developed or used, quizzes or other informal assessments employed, scoring guides you provided, graphs of student performance from pre-to-post assessments, etc.</td>
<td>A short - probably not more than two pages - written description of the development and/or use of the assessments procedures or tools you provided, what these told you about your students’ learning, and how you used the information you obtained.</td>
</tr>
<tr>
<td>Implementation</td>
<td>Samples of lesson plans and/or lessons implemented, examples of student work, notes on adjustments or modifications you made during the instruction, etc</td>
<td>A short paragraph around each lesson describing how the activities were implemented, how students received them, and any modifications you made and why.</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Graphs of student performance from pre-to-post assessments, copies of grading roster for the unit, examples of student work on assessment activities, other indicators of student progress.</td>
<td>A not more than three page written discussion of student progress and learning during the units. Include a re-examination of the teaching/learning context and its effects on the assessment results and instructional outcomes, an analysis of the assessment data for each student or groups of students within the class, a discussion, if appropriate, of the use you made of informal assessment procedures, and a discussion of your plans for re-teaching these units.</td>
</tr>
<tr>
<td>Reflection</td>
<td>A not more than three page written reflection on the overall teaching experience you have documented, the thinking, planning, and adjusting you did while developing and implementing the components of the instructional unit(s). When (if) you teach this unit again, how will you do it next time? What did you learn about yourself as a teacher and your students as learners from this sample of your teaching?</td>
<td></td>
</tr>
</tbody>
</table>
## ATTACHMENT B

### TEP II TEACHER EFFECTIVENESS PORTFOLIO RATING GUIDE: DIMENSIONS AND STANDARDS OF TEACHER PERFORMANCE

<table>
<thead>
<tr>
<th>DIMENSION</th>
<th>STANDARD$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Teaching</td>
<td><em>(In addition to the numerical description provided in the TEP II survey packet) this description includes a) aspects of community, district, and/or school, as well as b) aspects of the classroom—including students’ current pre-instructional status—that can influence teaching and learning, in terms of both instructional demand and support. In this section, thought has clearly been given to c) the implications of the teaching and learning context on curriculum, instruction, and assessment.</em></td>
</tr>
<tr>
<td>2. Planning/General (rationale underlying portfolio)</td>
<td>Description and discussion includes a) an overview of the instructional target(s) addressed in the portfolio, and b) an overview of the assessment system used in gauging whether these targets have been met. The discussion may also refer to c) students’ previous experiences, developmental levels, pre-instructional status as well as d) state, district, and community expectations (as appropriate). Essentially, this section should provide e) a coherent rationale for what will be found in the portfolio, and relates it to the learners present.</td>
</tr>
<tr>
<td>3. Planning/Specific (goals/objectives, lessons)</td>
<td>Discussion and/or artifacts demonstrate that the goals and objectives of this sequence of teaching and learning are a) clear, and likely are understandable to students and other teachers at the same level of schooling. The evidence also suggests that plans are b) developmentally appropriate, yet c) challenging for current performance levels of students. Importantly, instructional planning should d) make careful note of, and strive for, consistency with state and district content and performance standards, for the appropriate benchmark. (Note that this standard does not necessarily have to be met through the production of formal “lesson plans.”)</td>
</tr>
<tr>
<td>4. Implementation (instruction)</td>
<td>Description, discussion, and/or artifacts demonstrate that instructional activities a) are aligned with targeted learning, i.e., likely to result in the learning being sought, b) varied to potentially address different learning preferences, c) engage and challenge (directly, or through adaptations or accommodations) all students, d) provide practice opportunities for the skills students are expected to demonstrate, and e) are responsively altered/tailored “in-flight” to meet students’ needs.</td>
</tr>
<tr>
<td>5. Assessment</td>
<td>Description and/or artifacts demonstrate that assessments (whether formal or informal) are: a) sufficient to represent the learning targeted, b) aligned with targets for learning, c) have clear and understandable directions, items, and/or scoring procedures, d) possess characteristics likely to enhance trustworthiness, e) are feasible to administer and score, f) show a diversity of assessment approaches, and g) are developmentally appropriate for the students present.</td>
</tr>
<tr>
<td>6. Outcomes (description and analysis of student progress)</td>
<td>Description, analysis, and/or artifacts a) clarify the effects of the teaching/learning context on learning results, b) are grounded in assessment results (whether formal, informal, or a combination thereof), c) examine the relative performance of different groups as well as individual students, d) bring together formal and informal assessments for a fuller picture of learning, and thus, e) enhance the reader’s understanding of the learning presented by providing conclusions that are consistent with the results, tying results to the stated targets of learning, and providing a useful summary of learning.</td>
</tr>
<tr>
<td>7. Reflection</td>
<td>Discussion and/or artifacts demonstrate a “stepping back” from events or actions. This section is a) analytical and/or integrative of causative or intervening factors, findings, and perspectives and may recognize inconsistencies. This section b) speaks to lessons learned by the teacher, and plans for future development. This section c) goes beyond technical and practical emphasis on ends and means to also surface personal beliefs and make judgments about whether practice is equitable, just, and respectful of others.</td>
</tr>
<tr>
<td>II. Learning</td>
<td>The various sections of this teacher’s portfolio provide descriptions, discussion, and supporting evidence that, taken together, construct a credible case that non-trivial learning has occurred, in terms of both quality and quantity, over the period of instructional time represented in the portfolio.</td>
</tr>
<tr>
<td>III. Focus on Standards</td>
<td>Throughout the portfolio, there is overt and consistent attention to the Oregon content and performance standards for student learning that describe what, and to what degree, students are to know and be able to do, at sequential benchmarks in their K-12 school careers. In other words, it is apparent in looking at the contents of this portfolio that the Oregon standards have played a prominent role in determining planning, curriculum, instruction, and assessment.</td>
</tr>
</tbody>
</table>

---

$^1$ The levels of teacher performance reflected in these standards are thought appropriate for beginning teachers (1st, 2nd, or 3rd year teachers). The teaching standards have been derived from descriptions of acceptable levels of performance for each dimension contained in Teacher Standards and Practices Commission (TSPC) standards for granting initial license to teach in Oregon, and from the more detailed scoring guides for preservice teacher work samples contained in *A Guide to Teacher Work Sample Methodology*, published by the Teacher Effectiveness Project: Phase I, Teaching Research Division, Western Oregon University.
IV. Summary Rating

a) Each required dimension (section) of the portfolio is included, along with supporting documentation or artifacts; b) Each required section of the portfolio is addressed in a thoughtful, accessible, and substantive manner; c) The portfolio is cohesive, i.e., the body of the teacher's work represented in the portfolio fits together in a sensible way; d) The portfolio attends overtly and consistently to issues of classroom, school, and community contexts; e) The portfolio attends overtly and consistently to issues of learner diversity; f) The portfolio attends consistently to issues of alignment among state standards, planning, curriculum, instruction, and assessment.

Notes for Raters

1. It may well be that a teacher philosophically disagrees with a standards-based approach to public education, and states as much in their portfolio. This should not result in marking down any section, or the portfolio as a whole. At the same time, given that Oregon has adopted a system of schooling, and has defined success within that system by the formulation of content and performance standards, and that teachers are in some sense a part and agent of the state’s educational/social compact, then it seems reasonable to expect that teachers would attend to the standards in their planning, instruction, and assessment. That is, whatever the teacher’s philosophical dissonance, the portfolio should reflect a consonance with Oregon standards to receive high marks.

2. In arriving at section ratings, a “preponderance of evidence” rule is in effect. For example, the assessment standard contains seven (7) parts. If a teacher’s portfolio addresses all seven parts, the standard is met. If 4, 5, or 6 of 7 are addressed, the standard is partially met. If 3 or less are met, the standard is not met. Please note that each part of the standard is not itself rated met, partially met, or not met. Rather, the rater should simply determine if the part is there as described, or not.

3. To provide feedback to teachers (mainly in the context of performance evaluation, or evaluation for licensing), raters should note in the appropriate cell of the portfolio ratings matrix, what parts of a standard were not addressed by the portfolio maker. In this way, teachers will be able to identify areas that require improvement, and raters will be able to identify reasons for their ratings.
## TEP II TEACHER EFFECTIVENESS PORTFOLIO SCORING SHEET*

**Portfolio Owner:** ________________  **Portfolio Rater:** ________________  **Date:** ________________

### Levels of Performance

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setting / Context</td>
<td></td>
<td>c</td>
<td>a, b</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Planning General</td>
<td>b</td>
<td>c, e</td>
<td>a, d</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Planning Specific</td>
<td></td>
<td></td>
<td>a, b, c, d</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Implementation</td>
<td>e</td>
<td>d</td>
<td>a, b, c</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Assessment</td>
<td></td>
<td>c, d, e, f, g</td>
<td>a, b</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Outcomes</td>
<td></td>
<td>a,</td>
<td>b, c, d, e</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Reflection</td>
<td>a, c</td>
<td>b,</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>II. Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>III. Focus on Standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>IV. Summary Rating</td>
<td>d,</td>
<td>b,</td>
<td>a, c, e</td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

* Entries shown on the scoring sheet illustrate its use with the scoring standards outlined in ATTACHMENT B

**Rater Comments:**

---

15.15
A PROFILE OF TEACHING IN A STANDARDS-BASED CLASSROOM

This instrument is designed as an observation tool to assess the extent to which teachers engage in classroom practices consistent with “standards-based” teaching, including:

1. How the teacher communicates to students what is to be learned (learning outcomes desired);
2. How the teaching and learning activities align with and support the communicated outcomes;
3. How the teacher varies instructional activities, materials and procedures to support student in their learning;
4. How the teacher varies content, and engages students with content to help them in their learning;
5. How the teacher monitors student work to adapt instruction;
6. How the teacher provides feedback to students about their work;
7. How the teacher creates a classroom environment that supports student learning to high standards.

The instrument consists of two sections and is intended for use by a trained observer present in the classroom. Part one asks the observer to identify whether specific teacher behaviors were observed, and provides space to describe what the teacher actually did. Part two asks the observer to make a summative rating around each cluster of teacher behaviors observed.

The observation should span two entire instructional periods during a school day that have a clearly identifiable subject matter focus. These may be in the morning, in the afternoon, or one in the morning and one in the afternoon. A SEPARATE OBSERVATION FORM AND RATING PACKET, HOWEVER, NEEDS TO BE PREPARED FOR EACH OBSERVATION MADE.

Observer’s Name: Date of observation:

1st observation: From: To: 2nd observation: From: To:

Teacher’s name: Grade: # of students:

Content focus of observation 1: Content focus of observation 2:

Please provide a brief summary description of the content and instructional activities you observed during this time period. If additional space is needed, attach a separate page. Also circle the learning tasks listed below in which students were engaged during the class period observed, and under each task circled provide a rough estimate of the proportion of class time spent in each task.

Check one: 1st observation period ____ 2nd observation period ____

<table>
<thead>
<tr>
<th>remember</th>
<th>explain</th>
<th>apply</th>
<th>evaluate</th>
<th>solve</th>
<th>create</th>
</tr>
</thead>
</table>


TEACHER BEHAVIORS AND PRACTICES DIRECTLY INFLUENCING THE INTELLECTUAL ENGAGEMENT OF STUDENTS IN LEARNING

Aligning and varying content to support students in their learning

- effectively and accurately communicates content
- varies content to accommodate differences in learners
- connects content to real-life situations
- connects content across disciplines
- engages students in multiple-level and/or higher order thinking tasks involving content
- identifies and addresses misconceptions about content
- demonstrates command of subject matter knowledge within the lesson observed

Creating a classroom environment that supports students in their learning

- maintains a positive, learning focused classroom environment
- monitors and manages individual and group behavior to maximize learning for all students
- manages time and resources to maximize learning for all students
- quickly and effectively resolves conflicts between students if they arise
- creatively interweaves content, activities, and discourse to engage students in their learning
- students help other students in their learning, and help their teacher as needed or requested
- students engage in self-directed learning
- students ask clarifying questions
- physical features of the classroom provide an inviting and supportive context for learning