Step G. Further Evaluate Qualifying Measures for the Purpose of Measuring Growth

Those promising assessments that have been verified by the team as aligned to critical content and informative and useful to educators and administrators are candidates for further consideration for use as a DDM. During this second phase of evaluation, team members will concentrate specifically on determining whether an assessment can be used for the purpose of measuring student growth. The key question for evaluators during this step is as follows:

- Does this measure detect some type of change in performance that represents a gain that can be attributed to effective instruction?

Why is it so important that a DDM measure student growth? In a recent technical guide for Commonwealth stakeholders,1 ESE provided the following explanation:

Student growth scores provide greater insight into student learning than is possible through the sole use of single-point-in-time student achievement measures . . . Achievement scores provide valuable feedback to educators about student attainment against standards, but taken by themselves may not be a sufficient reflection of student progress.

In addition, DDMs that measure growth help to “even the playing field” for educators—allowing educators who teach students who start out behind to have a similar opportunity to demonstrate their impact on student learning as educators who teach students who start out ahead . . . In fact, measures of growth actually create opportunities to identify where low-achieving students are making tremendous gains. Therefore, assessments that measure growth allow schools and districts to rate educators based on their impact on student learning, and not on which students they happen to be teaching. (p. 11)

The goal for this step is to identify those promising assessments that (a) document what students know and can do before instruction and what they know and can do following a course of instruction or (b) capture important learning milestones reached throughout the course of instruction. An appropriate DDM must be useful for detecting a change in performance that can be interpreted as growth, because, by design, that change will be strongly linked to what is taught in a classroom or course. It should now be clear why the first

Key ESE-Developed Resources to Support Step G

- Technical Guide A (see especially pp. 25–28)
- Technical Guide B (see especially pp. 6–7 and 11–20 and Appendix B)
- Assessment Quality Tracking Tool
- Using Current Assessments in District-Determined Measures: Leveraging the Curriculum-Embedded Performance Assessments from the Model Curriculum Units (see especially pp. 2–3, 6–8, and 11)
- Webinar Series #2 (slides 21–41), #5 (slides 20–27 and 32), #7 (slides 11–12, 25–32, and 34), and #8 (slides 12–21 and 23–28)
- Implementation Brief: Scoring and Setting Parameters

---

1 Technical Guide B: Measuring Student Growth and Piloting District-Determined Measures.
criterion in Step F, alignment to content identified as central to instruction, is a prerequisite for the capacity to measure student growth.

**Approaches to Measuring Growth.** As discussed in the ESE resources cited as key for this step, a number of approaches to measuring growth are familiar to educators and do not require complex psychometric methods or statistical computations. Each of these approaches has unique strengths and limitations. Because of the unique characteristics of each CVTE program, one approach might work best in one program while another approach is better suited for a different program. In addition, many measures can be adapted to a variety of approaches to measuring growth, so decision-makers will want to use their professional judgment in weighing the pros and cons of each, considering competing goals, and determining the approaches best suited for their contexts.

Tables 4–7 describe four common approaches to measuring growth, the strengths and limitations of each, and specific considerations about each approach from ESE and from the research and measurement communities. Following each table is a summary of a CVTE program-submitted example of a measure using the growth approach described in the table; the full measures are included in Appendix A.

**Common Approaches to Measuring Growth: Pre-Test/Post-Test Assessment**

Description: A pre-test/post-test looks at the difference between scores on two assessments administered at different points in time. These assessments help educators understand how much a student has learned as a result of instruction. Districts interested in this option may use traditional or non-traditional measures as pre-test/post-test assessments.
**Table 4. Strengths and Limitations of Pre-Test/Post-Test Assessment**

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Approach is familiar to educators.</td>
<td>• Assumes a common scale that allows scores to be compared across different points in time and assumes score differences to be interpretable as representing increased mastery of specific knowledge or set of skills.</td>
</tr>
<tr>
<td>• Gain score (difference between the two scores) approach is easily computed but also can be the foundation for more complex analyses.</td>
<td>• May be ceiling effect if measures do not have sufficient number of very hard items or elements so that students with strong baseline scores can demonstrate growth.</td>
</tr>
<tr>
<td>• Can use identical measures administered twice or comparable versions.</td>
<td>• Raw gain score should not be used to rank students.</td>
</tr>
<tr>
<td>• Provides descriptive information about both direction (positive or negative) and magnitude (how much) of change.</td>
<td>• Items or elements on measures must be carefully developed or selected to ensure that change in what students know and can do is directly linked to what was taught.</td>
</tr>
<tr>
<td>• When used to examine magnitude of gain at the student level or average gain for a classroom, this approach is appropriate and can have sufficient precision.</td>
<td>• May create incentive to keep pre-test scores low.</td>
</tr>
<tr>
<td>• Growth expectations can be aligned to an absolute standard (student made progress toward a set goal) or norm-referenced (student made more or less progress than peers). If model of choice uses absolute standard, value judgments are necessary for low, moderate, and high growth. If model is norm-referenced, rating can be interpreted as a percentile ranking.</td>
<td></td>
</tr>
</tbody>
</table>

**Other Considerations**

- Can be reported as raw gain (e.g., 20-point increase from pre-test to post-test) or percentage increase (20 percent increase from pre-test to post-test). The method used should ensure comparability between students beginning at different levels of achievement. Districts may also use an approach where different sets of parameters are used to determine high, moderate, or low growth based on pre-test scores.

- If using comparable versions of a test, need to ensure that measures are not so different that changes in performance may be due to variations in the measures rather than to actual student growth.

- Time interval between tests must be explicit and intentional. It may be based on time (e.g., after ten weeks of instruction) or curriculum (e.g., after 25 lessons).

- If linked assessments are scored on a common scale, scores yielded should represent the full range of performance at the beginning and the end of school year.
Common Approaches to Measuring Growth: Repeated Measures Assessment

Description: With a repeated measures assessment, one or more assessments are administered multiple times to get information about how students performed at different points in time. This approach generalizes gain scores from multiple time points through the calculation of an average gain or slope (learning trajectory). Districts interested in this option may use traditional or non-traditional measures as repeated measures assessments.

Table 5. Strengths and Limitations of Repeated Measures Assessment

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Allows educators to use current assessments that are already administered multiple times during the school year.</td>
<td>• Requires authentic tasks so that improvement over time reflects growth rather than familiarity with the assessment.</td>
</tr>
<tr>
<td>• Generally relatively short and deeply integrated into the curriculum.</td>
<td>• Requires educator capacity to record and organize multiple data points for each student.</td>
</tr>
<tr>
<td>• Provides information to educators about improvements at various times across the year.</td>
<td>• Intervals between testing should be consistent across students and over time.</td>
</tr>
<tr>
<td>• Educators receive continuous feedback about students, and can modify their practices accordingly.</td>
<td>• Limitations of pre-test/post-test are relevant to this approach e.g., assumes a common scale that allows scores to be compared across different points in time and assumes score differences to be interpretable as representing increased mastery of specific knowledge or set of skills.</td>
</tr>
<tr>
<td>• Growth score can be generated using methods ranging from simple to sophisticated.</td>
<td></td>
</tr>
<tr>
<td>• Strengths of pre-test/post-test are relevant to this approach. However, precision of growth estimate generally increases using multiple time points.</td>
<td></td>
</tr>
</tbody>
</table>
Other Considerations

- Helpful to look at results graphically.
- Authentic performance tasks that (a) closely match the learning an educator wants students to demonstrate, and (b) are designed to show improvement over time, as students gain more sophisticated knowledge and/or improve skills, are the best candidates for use as a repeated measures assessment.
- For the purposes of determining high, moderate, or low growth, districts can compute a growth score by taking the difference between averages of several observations.

Lessons from the Field: Example of a Repeated Measures Assessment

CVTE Program: Carpentry

Assessment Name: Carpentry Competency Chart

Item Type: Competency-based performance task checklist

Description: This instrument measures technical skills attainment in carpentry using a performance task checklist. It may be used as a repeated measure to gauge student progress toward mastery of shop equipment safety training and shop tool use. Competencies in the use of the following tools and equipment are captured through educator observation of student demonstrations and task performance: circular saw, radial arm saw, jointer, planer, table saw, wide-belt sander, and slide-compound saw.

Note: The complete DDM is provided in Appendix A.

Common Approaches to Measuring Growth: Holistic Evaluation

Description: A holistic evaluation is based on a review of a collection of student work samples gathered systematically over a sustained period of time. Using a scoring rubric, growth is evaluated holistically across all work samples. Using a growth rubric, a rater evaluates the collection of student work holistically. This approach is typically associated with non-traditional measures such as a set of on-demand performance activities/demonstrations, portfolios, and observation-based measures (observing one or more behaviors or actions at key intervals and providing an overall rating).
Table 6. Strengths and Limitations of Holistic Evaluation

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creates an opportunity to evaluate students’ changing proficiency in performing complex tasks that have real-world relevance.</td>
<td>Can be time-consuming to administer. Requires time to observe a wide range of performances to allow for development of meaningful rubrics.</td>
</tr>
<tr>
<td>Scoring rubric can be designed to combine the processes of scoring student work, determining growth, and setting parameters for growth.</td>
<td>Standardized administration conditions (when? how many elements? who chooses the elements?) must be enforced.</td>
</tr>
<tr>
<td>Use of a growth rubric can address important validity concerns raised by scoring performance assessments individually and then computing a growth score from the individual scores.</td>
<td>Requires expertise to develop an effective holistic rubric designed explicitly for the purpose of measuring growth. Also challenging to set the criteria on which students will be judged and to define operationally these criteria in the form of a useful scoring rubric.</td>
</tr>
<tr>
<td>May be better able to show growth along a wide spectrum as it allows the scorer to consistently identify growth in student work that may have very different strengths and weaknesses. For example, it is useful when the students in a particular grade, subject, or course have quite varied levels of ability but are working on a common task.</td>
<td>Can be time-consuming to score. Also requires training and calibration to ensure consistent scoring against rubric (inter-rater reliability challenges).</td>
</tr>
<tr>
<td>Useful for measuring complex skills (e.g., reasoning), performing multi-step tasks, or analyzing work products that require considerable interpretation when being evaluated.</td>
<td>Must ensure that products and/or behaviors observed are directly linked to what was taught (curriculum embedded).</td>
</tr>
<tr>
<td>Rubrics are a communication tool. They make learning expectations and evaluation criteria explicit and can be used to help students get a clear vision of the task and the learning goals to which the task is linked.</td>
<td>Different performance levels should be linked systematically to the development of expertise. Educators sharing a common DDM should reach agreement on the types of evidence/behaviors associated with performance at different levels.</td>
</tr>
<tr>
<td>Illuminates where students started and how an educator helped them grow.</td>
<td>Increasing the number of elements or tasks can improve the reliability of scores obtained but also adds to the time and expense involved in administering and scoring.</td>
</tr>
</tbody>
</table>
Other Considerations

- Rubric should include detailed descriptions of what growth looks like at each performance level.
- Evaluation criteria should be specific enough to capture the most essential ingredients of the skill being measured but not so detailed as to create undue burden for scorers in terms of length. Criteria should help educators distinguish among students’ performances but not be overwhelming.
- Most effective when it reflects the intended targets of learning and clarifies the important dimensions of each target.
- Development of performance level descriptors will be an easier task if samples of actual student work are considered and those that best match various performance levels are identified. Then the characteristics of each can be captured in the descriptors.
- Acceptable levels of agreement between raters can be established through the use of anchor papers and other methods for calibrating scores.
- May be more efficient to borrow a strong growth rubric that already exists and adapt it for use in your desired context.
- To maintain scoring integrity, recommended practice is to arrange for educator peers to score each other’s sets of portfolios (or to use an alternative process that utilizes external raters).
Lessons from the Field: Example of a Holistic Evaluation

Tom Belland, Vocational Coordinator
Blackstone Valley Regional Vocational Technical High School

CVTE Programs: Automotive Technology, Carpentry, Health Assisting, Plumbing

Assessment Name: Certificate of Occupational Proficiency (C.O.P.)/Skills Assessments

Item Types: On-demand writing and performance tasks

Description: Massachusetts C.O.P./Skills Assessments were created to help schools gauge student progress on comprehensive exams in a competitive atmosphere. The assessments were created by working groups of CVTE educators and administrators from several schools in the Commonwealth. The four exams listed below are the start of a lengthy process that will encompass all state-recognized CVTE programs at the secondary level. They will also provide student and district data to those schools participating in a C.O.P./Skills pilot data cohort now in the development stage.

- **Automotive Technology**
  5 hands-on skills assessments
  Written exam: Workforce Ready System (ASE student certification, A.Y.E.S.)

- **Carpentry**
  1 hands-on skills assessment (highly comprehensive)
  Written exam: Workforce Ready System (Carpentry)

- **Health Assisting**
  5 hands-on skills assessments
  Written exam: Workforce Ready System (Nursing Assisting)

- **Plumbing**
  5 hands-on skills assessments
  Written exam: State sanctioned Tier test or Workforce Ready System (Plumbing)

Note: The complete DDM is provided in Appendix A.
Lessons from the Field: Example of a Holistic Evaluation

Leslie Weckesser, Vocational/Technical Director
Southeastern Regional Technical Vocational High School

CVTE Programs: Applicable in Multiple programs; widely adaptable
Assessment Name: Capstone Portfolio and Scoring Rubric 2014
Item Types: Work samples collected over time

Description: End-of-year, capstone project for high school students in grades 9-12. Students in grades 9 and 10 compile shop portfolios. Grades 11 and 12 students produce a portfolio and a senior project to integrate and apply knowledge, skills, and understanding in their chosen shop. Four project and presentation rubrics are included for each school year from freshman to senior. For instance, the sophomore rubric captures student performance data on four components completed during the academic year:

**Term 1:** Vocational My Access assignment on a new skill learned in the student’s permanent vocational program. Students are encouraged to include photos of the steps/process/completed project. Students must obtain a My Access Score of 5 or above.

**Term 2:** Vocational My Access assignment on a new skill learned in the student’s permanent vocational program. Students are encouraged to include photos of the steps/process/completed project. Students must obtain a My Access Score of 5 or above.

**Term 3:** Complete a job application with all necessary information. Students must save a completed copy for their portfolios.

**Term 4:** PowerPoint presentation on a new skill learned in the student’s permanent vocational program. Students are expected to give a 5-10 minute presentation about what they learned. Vocational educator can use discretion in assigning topic. Students must also provide evidence of searching and applying for a summer job and save a copy of application for their portfolios.

*Note:* The complete DDM is provided in Appendix A.
Common Approaches to Measuring Growth: Post-Test Only

Description: A post-test only is a single administration of a measure of what a student knows and can do following instruction. It requires collection of additional information to measure growth. A post-test only approach typically involves use of traditional direct measures such as summative exams (unit exam, mid-term or final exam, end-of-year exam, or end-of-course exam) or non-traditional measures such as capstone projects. It also may use indirect measures (e.g., number of students electing more advanced courses in relevant content areas in subsequent years).
Table 7. Strengths and Limitations of Post-Test Only

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Often the best approach when a measure of achievement or performance can only be administered once (e.g., building a house). Also useful when an existing measure is strongly linked to instruction (e.g., curriculum-embedded assessment) or is a well-established external tool (e.g., certification exam, Advanced Placement test).</td>
<td>• Developers will need to consider how to capture baseline information within the same year or course and produce scores that could be compared to the post-test.</td>
</tr>
<tr>
<td></td>
<td>• Districts need to think carefully about how they are measuring a student’s ability relative to where the student started.</td>
</tr>
<tr>
<td></td>
<td>• Single-year scores are not precise indicators of educator performance because results are confounded by cohort effect (i.e., different classes of students are systematically different for reasons not related to instruction).</td>
</tr>
</tbody>
</table>

**Other Considerations**

- Wherever possible, districts are encouraged to use other approaches to measuring growth.
- One approach to using a post-test only model would be to identify key knowledge or skills that are taught during the year or course and develop a baseline test/task or scoring method that allows educators to measure growth. The baseline test/task should be as similar as possible to the post-test in order to provide more confidence in interpreting growth results.
- It is critical to include additional information that helps predict the expected level of achievement, such as the difference from the previous unit’s test or the previous year’s graduation rate.
- A measure must be of exceptionally high quality in terms of content alignment and usefulness to justify this approach. DDM teams also will want to consider item quality, test length, accessibility for all students, and when and how the measure is administered and scored.

**Lessons from the Field: Example of a Post-Test-Only Assessment**

**Leanne Lyons**

**Whittier Regional Vocational Technical High School**

**CVTE Program:** Marketing

**Assessment Name:** General Industry Mid-Term Exam

**Item Types:** Selected response, constructed response, and performance checklist

**Description:** Marketing mid-term exam comprised of 100 selected-response items, 10 open-ended essay questions, and a performance task. Principles of marketing and customer service are assessed by this instrument. Multiple-choice items (#1-60) assess Strand 2 of the CVTE Frameworks, while items #61-100 assess Strands 1, 4, and 5.

**Note:** The complete DDM is provided in Appendix A.
**Final Decision-Making.** Overall, the DDM team will be looking for measures that are accessible to all students; that yield useful, meaningful results; and whose use can be justified to stakeholders in terms of benefit vs. cost. It is important for decision-makers to remember that the measures selected will benefit from review and refinement over time. As stated in a key ESE-developed resource,²

> When engaging in this work, do not let “perfect” be the enemy of “good.” Since districtwide implementation of many DDMs is new, it will take time to resolve every detail, such as nuanced questions about details of administration. The key focus is the use of measures that are meaningful and useful to educators at the classroom, school, and district levels. (p. 24)

CVTE educators and administrators have found WestEd’s *DDM Review and Discussion Guide* (see Appendix D-2) useful for determining the approach to growth that is associated with a particular assessment, evaluating an assessment’s usefulness as a measure of growth, and examining other considerations.

The anticipated outcome from Step G is a list of all assessments evaluated and decisions made about each in terms of its capacity to measure growth and its feasibility.

---

**Lessons from the Field: Measuring Growth**

**Tom Belland, Vocational Coordinator**

**Blackstone Valley Regional Vocational Technical High School**

Comments from Mr. Belland about measuring growth:

> Through regional meetings over the last two years, Blackstone has been training CVTE administrators and educators in the use of instructional units, assessments, and growth rubrics tied to student portfolios used to earn a Certificate of Occupational Proficiency (C.O.P.). We believe that, with some refinement, the C.O.P.s show real promise for use as DDMs. So far, we have taken the 17 C.O.P. packages we created for Blackstone Valley between 2011 and 2013, and worked with several nearby districts to customize five of them for use as DDMs. [See Appendix B for samples of the C.O.P. DDMs for Automotive Technology, Carpentry, Health Assisting, and Plumbing.]

As we have worked together to strengthen modified C.O.P.s for use as DDMs, we are reminding CVTE colleagues that their proposed DDMs do not necessarily need to measure growth over the whole school year. A measure that shows that a student met expectations for progress for a particularly important curriculum unit, for instance, could be considered evidence that the student is on track to achieving expected growth in one year’s time. In this way, the formative assessments that educators use to inform instruction during the school year also can potentially be used as DDMs. Recent research confirms the power of interim assessments to improve both instructional practice and student learning, bolstering the case for considering common unit, interim, and quarterly assessments as strong candidates for use as DDMs as long as pre-test or baseline data are available.

---

Lessons from the Field: Selecting the Most Promising Measures for Use as DDMs

Marcia Kessler, Technology Integration/Data Specialist
Old Colony Regional Vocational Technical High School

Marcia Kessler and Jackie Machamer, Vocational Coordinator, who are leading the DDM initiative at their school, have been working with academic and CVTE educators over the past four years to develop common assessments that integrate both academic content and technical skills. Ms. Kessler shared this insight:

Some would say that educators in Massachusetts suffer from “chronic implementation fatigue.” The new educator evaluation system requires educators to document evidence of their practice, and the DDM initiative for the Student Impact Rating is seen as another layer of additional evidence that they are required to collect. The DDM initiative has raised concerns and fears that educators will have to develop new assessments and will be judged using these new measures of student growth. However, educators and administrators should consider how they already know what their students can and cannot do and look at these indicators more closely to determine how their existing measures can be modified for use as student growth measure[s].
Lessons from the Field: Selecting the Most Promising Measures for Use as DDMs

Marcia Kessler, Technology Integration/Data Specialist
Old Colony Regional Vocational Technical High School

(Continued) While exploring assessment options and approaches for their district’s DDMs, Ms. Machamer and Ms. Kessler started with an existing assessment, a portfolio that was used as a summative measure, which they then modified for use as a repeated measure to demonstrate student growth. They are currently working with CVTE educators to develop a common rubric for the portfolio. They aim to modify the types of evidence collected and the time lines so that the portfolio can be used to monitor student growth at the beginning and end of each year and as a progression over four years as well. The portfolio will also be used with ELA educators as a repeated measure to reflect academic growth.

Ms. Machamer and Ms. Kessler are also working with educators at Old Colony to bridge academic content with technical skills through the development of a joint research project for grade 11 students. Grade 11 English educators and CVTE educators are developing a research paper that integrates ELA, technical skills, and employability skills; she is bringing together the English educators and CVTE educators to develop a common rubric to assess the research papers. The research papers will be used as a post-test-only performance assessment and will likely serve as one of the DDMs for the junior year.

Ms. Machamer and Ms. Kessler aim to assist CVTE educators in developing pre-test/post-test assessments for measuring growth in each program. They, as well as other members of the DDM team, plan to bring together educators in related career clusters to show them how to use the Edwin Teaching and Learning system to design pre-test/post-test assessments. Within Edwin, educators can easily select appropriate questions for their particular CVTE program, which are then grouped according to Bloom’s taxonomy to develop pre-test/post-test assessments. The competency tracking functionality in Edwin will also serve as a tool to assist CVTE educators in measuring student growth.

During the 2014 school year, the emphasis of educator professional development at Old Colony has been on curriculum development and DDMs. Mrs. Machamer and Mrs. Kessler have been working with both academic and CVTE educators to develop a curriculum guide that provides the key ideas and critical content aligned with the CVTE Frameworks so that students know what they should learn and be able to do in a given year of a program. The next step is to develop an assessment guide that will determine the best measures for the critical content. They aim for each educator to have at least one performance-based assessment and one traditional test for use as a DDM. The district has some very good measures in place, which will be modified for use as DDMs, that integrate academic and technical skills and that will hopefully serve as equitable and effective measures of student growth.