Strategies for Improving District-Determined Measures in the Academic Content Areas:

Guidance for Using the Repeated Measures Approach to Examine Student Growth

June 6, 2014
# Table of Contents

Introduction: Purpose of This Guidance Document i

Section I. Using the Repeated Measures Approach to Examine Student Growth with an Emerging DDM 1

Section II. Example Assessment #1, Using the Repeated Measures Approach to Examine Student Growth 3

Section III. Suggestions for Refining Example Assessment #1 to Ensure Usefulness as a DDM 7

Section IV. Example Assessment #2, Using the Repeated Measures Approach to Examine Student Growth 11

Section V. Suggestions for Refining Example Assessment #2 to Ensure Usefulness as a DDM 20

Appendix A: Key ESE-Developed Resources to Support Districts with Implementing DDMs that Effectively Measure Student Growth 21

Appendix B: Recommended External Resources on Measuring Student Growth: Guidance from the Research and Measurement Communities 24
Introduction: Purpose of This Guidance Document

Educators and administrators engaging in the selection and/or development of district-determined measures (DDMs) have been challenged with answering a key question about their emerging DDMs: *Does this measure detect some type of change in performance that can be attributed to student growth?* In a key resource developed to support districts with this challenge,¹ the Massachusetts Department of Elementary and Secondary Education (ESE) explains why it is so important that a DDM measure growth:

Students come to school each year with a wide range of prior academic achievement and therefore begin their next year of instruction with varying levels of readiness to access the curriculum, a situation that is beyond the control of the educator assigned to teach them. Measuring educators’ effectiveness solely by the achievement level of their students cannot account for these prior conditions. By comparison, measuring growth can help level the playing field. Improvement in student performance is a more meaningful and fair basis for determining the trends and patterns that will yield the educator’s rating of impact on student learning, growth, and achievement. (p. 8)

One strategy for supporting DDM developers in Commonwealth districts with this challenge is to provide concrete examples of open-source, locally developed assessments that are useful for the purpose of measuring student growth. The aim of sharing a sample of promising DDMs is to stimulate further discussion about how, with a few strategic changes, existing assessments may be adapted for use as DDMs. According to ESE, the strongest DDMs will emerge from districts that have engaged in systematic study of the appropriateness and usefulness of their assessments specifically for determining what students have learned from their educators during a course of instruction.

A number of approaches to measuring growth described in a key ESE-developed resource, *Technical Guide B*, do not require complex psychometric methods or statistical computations.² Each of these approaches has unique strengths and limitations. 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.

This document is intended to support those educators and administrators who are considering a *repeated measures* approach to examining growth with their locally developed assessments.

---

¹ Massachusetts Model System for Educator Evaluation, Part VII: Rating Educator Impact on Student Learning Using District-Determined Measures of Student Learning, Growth, and Achievement.

² Companion documents have been developed that highlight three other common approaches to measuring growth: (a) a *pre-test/post-test* approach; (b) a *holistic evaluation* approach; and (c) a *post-test only* approach.
It includes the following:

- key resources developed by ESE that offer guidance for monitoring the technical quality of DDMs and for selecting a reasonable approach to measuring growth;
- a summary of the strengths and limitations of the repeated measures approach to examining student growth;
- sample assessments submitted by Commonwealth districts, with suggestions for how the developers might refine each measure for use as a DDM; and
- information about a number of external resources that educators and administrators may find helpful if they elect to pursue a repeated measures approach.

It is hoped that this guidance document will help district personnel transform strong locally developed measures of achievement and/or performance into promising DDMs that are particularly effective in evaluating student growth.
Section I. Using the Repeated Measures Approach to Examine Student Growth with an Emerging DDM

With a repeated measures approach, one or more assessments are administered multiple times to produce information about how students perform at key points in time. This approach generalizes changes in performance such as gain scores from multiple time points through the calculation of an average gain or slope (learning trajectory). This approach can be used with traditional tests of achievement (e.g., end-of-grade/end-of-course assessments or teacher-developed exams) or more non-traditional measures that are performance-, project-, or portfolio-based. This approach has a number of strengths and limitations, based on guidance from experts in the educational research and measurement communities. These are summarized in Table 1.\(^3\)

---

\(^3\) See Appendix B for research- and measurement-based resources for the repeated measures approach.
Table 1. Strengths and Limitations of the Repeated Measures Approach to Examine Student Growth

<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>• Repeated measures are 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 throughout the year.</td>
<td>• Limitations of pre-test/post-test approach are relevant to this approach—e.g., assumes a common scale that allows scores to be compared across different points in time, and allows score differences to be interpretable as representing increased mastery of specific knowledge or set of skills.</td>
</tr>
<tr>
<td>• Educators receive continuous feedback about students, and can modify their practices accordingly.</td>
<td></td>
</tr>
<tr>
<td>• Growth scores can be generated using methods ranging from simple to sophisticated.</td>
<td></td>
</tr>
<tr>
<td>• Strengths of pre-test/post-test approach are relevant to this approach. However, precision of growth estimate generally increases using multiple time points.</td>
<td></td>
</tr>
</tbody>
</table>

Other Considerations

Districts interested in using the repeated measures approach also may want to attend to the following considerations:

• Looking at results graphically can be informative.

• The best candidates for use as a repeated measures assessment are authentic performance tasks that (a) closely match the learning that 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.

• For the purposes of determining high, moderate, or low growth, districts can compute a growth score by taking the difference between the averages of several observations.
Section II. Example Assessment #1, Using the Repeated Measures Approach to Examine Student Growth

The following assessment was submitted by a Commonwealth educator who adapted an existing measure for use as a DDM. For more information about the assessment, please contact Kate Ciulla at: kciulla@scit.org.

<table>
<thead>
<tr>
<th>Locally Developed Assessment Well Suited for Use as a DDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submitted by: Kate Ciulla, World Language Department Head, Scituate Public Schools</td>
</tr>
</tbody>
</table>

**Course/Grade:** World Language/French and Spanish, Grades 7–8

**Assessment Name:** Common Oral Assessment for 7th and 8th Grades

**Item Type:** Performance-Based Oral Assessment

**Description:** This assessment packet includes two grade-specific oral assessments: one that can be used for students in grade 7 French or Spanish and one that can be used for students in grade 8 French or Spanish. Each oral assessment provides a prompt, to which students are asked to respond orally, answering specific questions or following directives during their presentations. The purpose of the assessment is to detect changes in oral fluency and in use of grade- and curriculum-specific vocabulary over the course of the school year.

Students are shown a picture and provided with a list of questions or points of discussion that they are expected to address during their oral presentation. They are allowed five minutes to plan their responses, and then their responses are recorded. Some accommodations are allowed for students with special assessment needs. Administrators document the length of each student’s presentation (number of seconds he/she is speaking), then collect the recordings and score them at a later date.

Students record several “practice” oral assessments during the year, and then complete a final assessment at the end of the year. Prior to the final performance, recordings are used to help educators set performance targets for subsequent units of instruction and determine areas in which students need individual support to increase oral productivity. Educators can use the collection of responses to determine growth over time and to assign final, end-of-year grades for oral production.

Instructions for administering and scoring the assessments are included in the assessment package. The scoring rubric guides educators with evaluating students on four criteria: organization and understanding of the topic, control of language, delivery and pronunciation, and vocabulary. For each criterion, students receive a rating of Unsatisfactory, Needs Improvement, Proficient, or Advanced. Performance level descriptors for each criterion at each level are intended to promote consistent and accurate scoring of student responses. Students receive points for each criterion, and then an overall point total is calculated.
Instructions for Administering Common Oral Assessment, Grades 7 and 8

a. Individually, students will be given 5 minutes to record their ideas pertaining to the picture series. They may write down vocabulary, verbs, complete sentences, etc.

b. They may not use a book, notes, or electronic devices during the brainstorm.

c. At the completion of 5 minutes, the teachers will inform the class that the recordings/presentations will begin.

d. Students with IEPs and 504s that indicate the use of a word bank will receive a vocabulary list pertaining to the picture series. They may also use their notes from the brainstorm to complete the assessment.

Instrument: Common Oral Assessment, 7th Grade

Prompt: You are one of the students in the picture below. Introduce yourself to the teacher in the picture. Include the following information in your introduction...

a. Greet the teacher
b. Ask how she is
c. Say how you are
d. State your name
e. Spell your name
f. Give your email address
g. State your age
h. Say goodbye
Instrument: Common Oral Assessment, 8th Grade

Prompt: You are dining at home with your host family in Madrid or Paris. Discuss your experience while referring to the image below. You may also use your imagination. Include the following...

- Describe what you are eating or what you want to eat.
- Describe what others in the picture are eating or what they want to eat.
- Tell what you think of the food.
- Offer the food to the others at the table.
Common Oral Assessment: 7th and 8th Grade

RUBRIC—ORAL PRODUCTION
Adapted from the Gates World Language Oral Production Rubric

Assignment ____________________________________________
Name ________________________________________________

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Performance Levels: Advanced</th>
<th>Performance Levels: Proficient</th>
<th>Performance Levels: Needs Improvement</th>
<th>Performance Levels: Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORGANIZATION and UNDERSTANDING OF THE TOPIC</td>
<td>• Clear consistent unity and sequence and/or • In-depth, insightful coverage of topic</td>
<td>• Fairly clear, logical sequence and/or • Good understanding + coverage of topic</td>
<td>• Some evidence of sequence and/or • Adequate coverage of topic</td>
<td>• Lack of unity and sequence and/or • Little or no understanding of topic</td>
</tr>
<tr>
<td>ORGANIZATION and UNDERSTANDING OF THE TOPIC</td>
<td>5 points</td>
<td>4 points</td>
<td>3 points</td>
<td>1 point</td>
</tr>
<tr>
<td>CONTROL OF LANGUAGE</td>
<td>• Very accurate; smooth; no patterns of error</td>
<td>• Accurate; occasional pauses; few errors</td>
<td>• Fair; frequent pauses; some significant errors</td>
<td>• Weak; no flow; repeated errors</td>
</tr>
<tr>
<td>CONTROL OF LANGUAGE</td>
<td>10 points</td>
<td>8 points</td>
<td>6 points</td>
<td>2 points</td>
</tr>
<tr>
<td>DELIVERY and PRONUNCIATION</td>
<td>• Communication is exciting and engaging and • No noticeable errors in pronunciation</td>
<td>• Communication is effective and interesting and/or • 1-2 slight errors in pronunciation</td>
<td>• Communication is mechanical in nature and/or • Repeated patterns of error in pronunciation</td>
<td>• Delivery amounts to little more than reading and/or • Pronunciation renders communication incomprehensible</td>
</tr>
<tr>
<td>DELIVERY and PRONUNCIATION</td>
<td>15 points</td>
<td>12 points</td>
<td>9 points</td>
<td>3 points</td>
</tr>
<tr>
<td>VOCABULARY</td>
<td>• Precise, rich and varied</td>
<td>• Appropriate; not repetitious</td>
<td>• Limited and repetitious</td>
<td>• Inappropriate</td>
</tr>
<tr>
<td>VOCABULARY</td>
<td>10 points</td>
<td>8 points</td>
<td>6 points</td>
<td>2 points</td>
</tr>
<tr>
<td>REQUIRED COMPONENTS</td>
<td>• Includes all</td>
<td>• Includes most</td>
<td>• Missing several</td>
<td>• Includes one</td>
</tr>
<tr>
<td>REQUIRED COMPONENTS</td>
<td>10 points</td>
<td>8 points</td>
<td>6 points</td>
<td>2 points</td>
</tr>
<tr>
<td>Point Sub-Totals</td>
<td>50 pts</td>
<td>40 pts</td>
<td>30 pts</td>
<td>10 pts</td>
</tr>
</tbody>
</table>

Total Number of Points (max 50, assuming Advanced on all criteria) = __________

Length of utterance = __________ seconds
This promising DDM is unique in that it is a measure of students’ oral production skills that can be administered at key time points during the school year. The assessment activity is closely linked to instruction during those time periods, and can easily be adapted to ensure ongoing alignment to valued content. It provides a common performance assessment for educators teaching Spanish or French in two middle school grades.

Suggestions for refining this assessment are shared to illustrate some of the possible ways in which districts might, with slight modification, use existing assessments as DDMs.

**Suggestion 1: Use similar but not identical prompts or stimuli.** With the repeated measures approach, it is important that the assessments measure change in student responses over time that are due to instruction and not due to familiarity with the prompt or other stimuli. Districts interested in using this approach may want to develop unique prompts for each testing event that measure the knowledge, skills, or abilities of students in a similar way—e.g., by asking them to describe a picture—but differ in terms of the context or the activity, people, or scene presented. Doing so will create an authentic assessment process while minimizing the likelihood of a possible overexposure effect.

With the current assessment, for example, developers might collect pictures of people engaged in a wide variety of activities that will encourage students’ oral production as intended. These might include pictures of students engaged in activities such as playing soccer, attending a concert, preparing a family meal, or visiting someone at the hospital. The goal will be to find stimuli that are interesting to the students and that target use of specific skills and/or vocabulary words learned during the period of instruction preceding the testing event.

**Suggestion 2: Standardize the administration schedule.** Collecting information about what students know and can do at multiple (three or more) time points strengthens the accuracy and trustworthiness of educators’ decisions about student growth, while providing stakeholders with real-time feedback on student learning. At a minimum, districts using the repeated measures approach will want to collect performance or achievement data at the beginning of the year, at mid-term, and again at the end of the year.

Districts considering this option will want to establish a standardized administration schedule so that all educators in a particular content area (e.g., Spanish) who are using the assessment collect recordings during specific time periods. If grade 7 world language educators are involved in this process, they can identify the critical milestone activities in their instruction that would be ideal time periods for data collection. Administering the assessment during those times would provide educators with feedback on the degree to which their instruction has resulted in the learning outcomes they intended. This information can be used formatively to adjust
instructional methods as needed and to guide decision-making about additional support for students who appear to be falling behind.

**Suggestion 3: Calculate growth using data from all time points.** Educators can use a number of different strategies to arrive at a rating for growth for each student. For example, using the assessment rubric provided on page 6 of this report, they would first note that the range of points possible for each criterion differs, as shown in Table 2.

**Table 2. Oral Production Rubric: Possible Range of Scores, by Criterion**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Performance Levels: Advanced</th>
<th>Performance Levels: Proficient</th>
<th>Performance Levels: Needs Improvement</th>
<th>Performance Levels: Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization and Understanding of the Topic</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Control of Language</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Delivery and Pronunciation</td>
<td>15</td>
<td>12</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Required Components</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

The educators who developed this assessment intended for these ranges to be different, in order to emphasize those criteria that they wanted to carry relatively heavier weight in the overall score. For example, Delivery and Pronunciation carries the greatest weight, relative to the other criteria. Similarly, performance at different levels carries point values that reflect the developers’ intent; students rated as Advanced across all criteria can earn up to 50 points, while students rated as Needing Improvement across all criteria can earn only 30 points.

Assuming that the assessment is administered five times, educators would assign point values, by criterion, for each student for each assessment event. They might then create a graph, such as the one shown in Figure 1, that displays the scores for a particular student on tests administered throughout the school year.
Educators may find this data display quite useful for instructional planning. At a glance, they can see that, in terms of Organization and Understanding of the Topic, Student A moved from Unsatisfactory on Test Event 1 to Needs Improvement on Test Event 2, and by Test Event 4 he had reached Proficient. His performance on Delivery and Pronunciation was more stable, with ratings of Needs Improvement on the first three test events and a rating of Proficient on the two subsequent testing events. Educators can use these data to monitor student growth between events and over the course of the school year, paying special attention to those criteria on which all students seem to be experiencing less-than-optimal growth, as they may suggest possible areas for instructional adjustment.

Using this information, a total gain score for each student can be calculated by summing the gains in scores between Test Events 1 and 5 for each criterion. As shown in Figure 1, between Test Events 1 and 5, Student A gained 3 points (4 minus 1) on Organization and Understanding of the Topic, 6 points (8 minus 2) on Control of Language, 3 points (12 minus 9) on Delivery and Pronunciation, 8 points (10 minus 2) on Vocabulary, and 4 points (10 minus 6) on Required Components. Overall, Student A’s growth score is 24 (3 + 6 + 3 + 8 + 4).

These scores can be used to separate students into groups of low, moderate, and high growth. A district using this approach could look at the distribution of all gain scores, from lowest to highest. Since the maximum possible gain is 40 points (a maximum of 50 points if scores for all criteria on Test Event 5 are Advanced, minus a maximum of 10 points if scores for all criteria on Test Event 1 are Unsatisfactory), educators might conclude that a gain score between 15 and 25 is a reasonable range for moderate growth, and thus set 0–14 as the range for low growth and 26–40 as the range for high growth. Using the example in Figure 1, with a total score of 24,
Student A would fall into the moderate-growth category. To confirm the reasonableness of these assignments, district personnel might want to review the distribution of student grades for that grade or course, to see if students’ grades are consistent with their growth scores. Districts using this approach will want to revisit the impact of their decisions each year and make adjustments as needed to ensure that no unintended negative consequences from test use are emerging.

In addition, a district might want to look carefully at student gains across test events for each criterion. For example, as shown in Figure 1, across the first two assessment events, for Control of Language, Delivery and Pronunciation and Required Components, Student A showed no gain between Test Events 1 and 2; this trend continued through Test Event 3 for Delivery and Pronunciation. Questions that districts and educators may ask in this circumstance include: Was this pattern typical for all students? What was the range of all gain scores for that criterion? Does this information seem consistent with other indicators of student learning that are related to that criterion? Educators may find this information useful for instructional planning and in checking their assumptions about how much students are learning between assessment events.
A second assessment submitted by a Commonwealth educator is presented on the following pages. For more information about the assessment, please contact Jamie Vitonis at jvitonis@slrsd.org.

Locally Developed Assessment Well Suited for Use as a DDM
Submitted by: Jamie Vitonis, ELA Coordinator, Silver Lake Regional School District

Course/Grade: English Language Arts (ELA), Grade 9
Assessment Name: Writing to Text
Item Type: Constructed Response

Description: This assessment packet includes three writing prompts that require students to develop effective arguments and find evidence in selected texts to support their positions. Students are also expected to draw upon their personal background experiences and/or refer to historical or current events in their essays.

The prompts are briefly summarized as follows:

- **Is there a downside to having high self-esteem?**
- **Is punishment the most effective way to change negative behavior?**
- **What makes a hero? Can heroes be ordinary or must they overcome a tragedy or major adversity?**

The prompts can be administered as stand-alone assessments at key time points in the school year, such as at the beginning of the year, at mid-term, and at the end of the year. Even though the contents of the prompts are different, they all require students to demonstrate the same set of skills. Educators can use student responses to better understand the ways in which students have grown in terms of (a) use of claims or taking a position, (b) organization, (c) claim development and use of evidence, (d) identifying opposing arguments, and (e) use of appropriate language.

In addition to each prompt, students are provided with a short passage that is intended to help them take a position on an issue and develop an effective argument to support their position, using evidence from the passage text. Students must understand the passages, including their rhetorical contexts, and then locate details in the passages that strengthen their arguments. These passages may be excerpts from nonfiction essays; articles from journals, magazines, or newspapers; or reprinted blogs or other web-based entries.

This assessment is aligned to the Massachusetts Curriculum Framework for grades 9–10 in both writing and reading informational text. It can easily be adapted for use in any ELA or social studies classroom.
Prompt #1

Is there a downside to having high self-esteem? Write an argument essay in which you defend your claim. You must use supporting evidence from two of the three sources. You must also include a reference to one of the following: personal experience, literature, history or current events.


All true champions have inner arrogance. It’s that confident feeling they have that says, “I have the edge over you,” or, “You don’t have a chance against me.” It’s the same feeling that Lou Brock had standing squarely on the bag after stealing yet another base--as if it was his alone. It’s the same feeling Muhammad Ali had as he prepared to prove once again that he was “the greatest.”

This confident attitude of “I can’t lose” is a powerful psychological tool great athletes use in gaining the edge over their opponents. They know there’s plenty of time to be reserved and humble --but the period spent getting ready to perform is not one of them.

Before a game, you should remind yourself of the many long hours of work you’ve put into your sport, the sacrifices you’ve made, the skills you’ve learned and the many successes you’ve enjoyed. You know you’re good. Soon you’ll be on center stage again to prove it. It’s a feeling of confidence only champions have the right to experience.


Self-esteem is how you feel about yourself. You may have high self-esteem—you believe you are a good person—or it may be low, and you question how “good” or “worthy” you are.

Everyone has low self-esteem at times. It may happen when someone says something bad about you or questions how well you do something. But if you often feel bad about yourself, you may stop believing in yourself. You may find it hard to meet your goals and enjoy life.

People who have depression or anxiety or who are sick or have a disability may have low self-esteem.

Building self-esteem is a lifelong process, but it can help you feel better about yourself and your life. Here are some ways to build your self-esteem. Start by picking something you’d like to try now. Later you can decide if you’d like to try other ways.
Think about yourself

Try to do things that keep you healthy, that show you respect yourself, and that give you pleasure.
- Eat a balanced diet, and be active.
- Wear clothing you like and feel good in.
- Make your room, apartment, or home special for you.
- Help someone or do a favor for somebody else.
- Take time to do something you enjoy.
- Spend time with people who like you and make you feel good about yourself.


Laura Rovi was smart enough to be lazy. An honor student at Elmhurst's York High School, she was accustomed to getting an A even when she cruised through a class.

She expected nothing less when she took a government course her sophomore year and let a classmate do all the work on their final group project, an advocacy video warning of the dangers of eating disorders.

This time, though, her lack of effort earned her a C — a mark that produced a curious reaction.

She wasn't guilty. She wasn't depressed. She was insulted.

"This was just in my face," Rovi, 18, recalled recently. "I was not used to that."

Rovi belongs to a generation of teens for whom praise has often come as readily as oxygen. They've been bathed from the cradle in affirmations and awards meant to boost their self-esteem — and, by extension, their prospects in life.

But some who research the psychology of teens have concluded that this trend, born of good intentions in the Age of Aquarius, has had toxic effects.

By their estimation, today's young people have been praised so much that some flail at their first taste of criticism or failure. Others develop a keen sense of privilege, believing they'll coast into a golden future regardless of their actual talents, accomplishments or willingness to work.

"There has been a pretty big shift in expectations. Adjusting to reality is going to be different," said Jean Twenge, a San Diego State University psychology professor whose research has found soaring teen self-esteem.

Twenge's conclusion is not universally accepted — other researchers have found no significant changes in self-esteem from previous generations — but it rings true in many schools and homes. And it has some adults asking themselves hard questions.

"It's this entitlement that is driving many of us crazy. It's like, where did we go wrong?" said Rita Berger, a West Chicago mother of a teenage son and daughter. "We're kind of the root problem. In our attempt to give (this generation) everything, they have not learned to work or appreciate things."
The self-esteem movement grew out of the work of therapists like Nathaniel Branden, who in the late 1960s wrote that internal negativity could lead to lack of achievement. Change what people think of themselves, he contended, and you can change their destiny.

**Prompt #2**

Is punishment the most effective way to change negative behavior? Write an argument essay in which you defend your claim. You must use supporting evidence from two of the three sources. You must also reference at least one of the following: personal experience, literature, history or current events.


There’s no such thing as a good day for a prisoner at the highest level of security within the Ohio State Penitentiary, a 504-bed supermax prison in Youngstown, Ohio. Every inmate lives alone in a 7-ft. by 14-ft. cell that resembles nothing so much as a large, concrete closet, equipped with a sink, a toilet, a desk and a molded stool and sleep platform covered by a thin mattress. The solid metal door is outfitted with strips around the sides and bottom, muffling conversation with inmates in adjacent cells. Three times a day, a tray of food is delivered and is eaten alone. The prisoner may spend 23 hours a day in lockdown, emerging to exercise once a day. The lights in the cell never go off, although they may be dimmed a bit at night.

If there’s not much to like about the conditions in Youngstown, there’s not much to like about the people confined there either. These are the men corrections folks like to call "the worst of the worst," the kind of felons who dealt drugs or led gangs or killed on the outside and continued to do so in prison. For them, maximum security would not be enough--only supermax would do. And say what you will about the draconian environment, it keeps them under control.

But that level of control may be counterproductive. It’s possible that the very steps we’re taking to keep society safe and such prisoners in check are achieving just the opposite. The U.S. holds about 2 million people under lock and key, and 20,000 of them are confined in the 31 supermaxes operated by the states and the Federal Government. That may represent only 1% of the inmate population, but it's a volatile 1%. Push any punishment too far and mental breakdown--or at least a claim of mental breakdown--is sure to follow. When that happens, a constitutional challenge can’t be far behind.


The seagulls begin squawking at 6 in the morning and the cigarettes cost too much, but Lars, 41, knows there are worse places to call home. On Bastoy, an island 46 miles (74 km) south of Oslo, he and 124 other residents live in brightly colored wooden chalets, spread over one square mile of forest and gently sloping hills. Besides enjoying views of the surrounding fjord, they go horseback riding and throw barbecues, and have access to a movie theater, tanning bed and, during winter, two ski jumps. Despite all its trappings, Bastoy island isn’t an exclusive resort: it’s a prison.
Arne Kvernvik Nilsen, Bastoy's governor and a practicing psychotherapist, describes it as the world's first human-ecological prison — a place where inmates learn to take responsibility for their actions by caring for the environment. Prisoners grow their own organic vegetables, turn their garbage into compost and tend to chickens, cows, horses and sheep. They also operate the ferry that shuttles a number of them to school and jobs on the mainland, make their own dinner (they're allowed to use knives) and chop wood (using axes and chainsaws). Although authorities carry out routine drug tests, the prison generally emphasizes trust and self-regulation: Bastoy has no fences, the windows have no bars, and only five guards remain on the island after 3 p.m. and on weekends. "They are among the worst criminals in Norway," says Nilsen. "But they keep the whole society alive and running."

In an age when countries from Britain to the U.S. cope with exploding prison populations by building ever larger — and, many would say, ever harsher — prisons, Bastoy seems like an unorthodox, even bizarre, departure. But Norwegians see the island as the embodiment of their country’s long-standing penal philosophy: that traditional, repressive prisons do not work, and that treating prisoners humanely boosts their chances of reintegrating into society. "People in other countries say that what Norway does is wrong," says Lars, who is serving a 16-year sentence for serious drug offenses. "But why does Norway have the world’s lowest murder rate? Maybe we're doing something that really works."

Countries track recidivism rates differently, but even an imperfect comparison suggests that Norway's system produces overwhelmingly positive results. Within two years of their release, 20% of Norway’s prisoners end up back in jail. In the U.K. and the U.S., the figure hovers between 50% and 60%. Of course, Norway’s low level of criminality gives it a massive advantage. Its prison roll lists a mere 3,300 inmates, a rate of 70 per 100,000 people, compared with 2.3 million in the U.S., or 753 per 100,000 — the highest rate in the world.


At Brandt and Audra’s house, beds need to be made before leaving for school. If you fail to make your bed, you go to bed thirty minutes early. The last time that Brandt failed to make his bed was three weeks ago. Audra fails to make her bed about four mornings a week. She goes to bed early each time.

Most parents believe that going to bed early is a good punishment. This seems like a good plan, but look at what is actually happening. This form of punishment works well for Brandt; he avoids the punishment by remembering to make his bed. He has decided that staying up a little later is important.

Going to bed early is not a punishment for Audra. She does not avoid it. Going to bed early has had no effect on her behavior; she is not making her bed. Maybe she likes going to bed early—I certainly do! Another punishment should be used for Audra, something that will change her behavior.

Punishment is a negative consequence. When used properly, punishment eliminates or reduces misbehavior. Using punishment correctly is difficult. It requires consistent follow-through. Too much punishment is harmful; it creates unpleasant feelings and drains energy. Punishment works,
but it is not easy to use effectively. Most parents believe that punishing a misbehavior will stop the child from repeating the misbehavior.

---

**Prompt #3**

*What makes a hero? Can heroes be ordinary or must they overcome a tragedy or major adversity? Write an argument essay in which you defend your claim. You must use supporting evidence from 2 of the three sources. You must also reference at least one of the following: personal experience, literature, history or current events.*

---


“We are doing this because the American fans are just beginning to understand that a sport, to be a real sport, has got to be contested on the basis of the best man or team winning--and the best has got nothing to do with how much brown or red or yellow tint is in a man's skin.”

During his baseball career, Jackie Robinson won many honors. He broke into the National League as the Rookie-of-the-Year, won the Most Valuable Player award and achieved a .311 lifetime batting average.

But none of those successes was greater than the courage he showed in breaking baseball’s color line by becoming the first black to play in the major leagues.

Robinson made a deal with Branch Rickey, general manager of the Brooklyn Dodgers, when he was signed in 1945 that he would never fight, even if provoked. Rickey knew many prejudiced ballplayers would not make it easy for Robinson, and he wanted to make sure Robinson would be able to withstand the pressure.

“Mr. Rickey,” Robinson asked, “you want a coward on your ball team?”

Rickey shook his head.

“I've got two cheeks; is that it?” Robinson asked.

Rickey nodded.

When he was verbally abused or when players slid into second base with their spikes high, Robinson kept his temper in check. He turned away, then got his revenge with a base hit or a diving catch. It is easy to fight back, but it takes real courage to turn the other cheek and show your class.

---


You've shown us, Boston, that in the face of evil, Americans will lift up what's good. In the face of cruelty, we will choose compassion. In the face of those who would visit death upon innocents, we will choose to save and to comfort and to heal. We'll choose friendship. We'll choose love.
Scripture teaches us, “God has not given us a spirit of fear and timidity, but of power, love, and self-discipline.” And that’s the spirit you’ve displayed in recent days.

When doctors and nurses, police and firefighters and EMTs and Guardsmen run towards explosions to treat the wounded -- that’s discipline.

When exhausted runners, including our troops and veterans -- who never expected to see such carnage on the streets back home -- become first responders themselves, tending to the injured -- that’s real power.

When Bostonians carry victims in their arms, deliver water and blankets, line up to give blood, open their homes to total strangers, give them rides back to reunite with their families -- that’s love.

That’s the message we send to those who carried this out and anyone who would do harm to our people. Yes, we will find you. And, yes, you will face justice. We will find you. We will hold you accountable. But more than that; our fidelity to our way of life -- to our free and open society -- will only grow stronger. For God has not given us a spirit of fear and timidity, but one of power and love and self-discipline.

And that’s what you’ve taught us, Boston. That’s what you’ve reminded us -- to push on. To persevere. To not grow weary. To not get faint. Even when it hurts. Even when our heart aches. We summon the strength that maybe we didn’t even know we had, and we carry on. We finish the race. We finish the race.

And we do that because of who we are. And we do that because we know that somewhere around the bend a stranger has a cup of water. Around the bend, somebody is there to boost our spirits. On that toughest mile, just when we think that we’ve hit a wall, someone will be there to cheer us on and pick us up if we fall. We know that.


Tom died the Saturday before Christmas. I don’t know much about Tom. He was about 60 and his death was unexpected. I don’t know his last name, because you don’t usually exchange last names when you do the sort of thing Tom was doing a few hours before he died.

The doctors said Tom overexerted and had a massive heart attack. I don’t know about that, I only know that a few hours before Tom died he was delivering a pickup truck’s worth of toys to children in a homeless shelter. Whatever the doctors say, it appears Tom’s heart worked beautifully right up to the last beat.

Last year I started a tradition for my column, pledging to devote the last Wednesday of the year to unknown or forgotten greatness around us. Folks who aren’t sports stars, superstars or rock stars — people who just do what matters to them and this makes other lives more special — people like Tom.

In February I read in the *Bangor Daily News* about Newport Police Chief Leonard Macdaid. If you do much driving in these parts, you know one thing about Newport: Don’t speed. Quite frankly, that’s the only thing I’ve ever heard of the Newport Police Department’s reputation.
Then I read about Chief Macdaid. The chief felt that keeping his community safe meant more than law enforcement: It included home visits to the elderly. A chance to check in, to chat — as the paper said — to make sure seniors are OK. “Is the house warm? Does everyone look well? Is there evidence of cooking or adequate groceries? Is the driveway cleared and the step ice-free?”

After I first read the story I called his office thinking it might be neat to devote a whole column to his sort of “justice” — the kind that truly protects the most vulnerable in his community. He was polite and cordial on the phone. But after talking to him a few minutes I realized he wasn’t the type of man you wrote about in a political OpEd of a newspaper. He was too humble, too unassuming, too genuine to have his name dropped anytime but now.

There is one thing the chief said that I would like to take exception to: that the elderly he visits are “by far the best generation this country ever had.” It can’t be. The chief isn’t in that generation, and his actions prove that every generation has the potential for greatness.

**Writing to Text, Grade 9: Argument Scoring Rubric**

<table>
<thead>
<tr>
<th>Item</th>
<th>0–2 points</th>
<th>3–4 points</th>
<th>5 points</th>
<th>6–7 points</th>
<th>8–9 points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overarching Claim/ Position</strong></td>
<td>Claim does not adequately address the prompt.</td>
<td>Claim may not be clear and/or may not fully address the prompt.</td>
<td>Addresses the prompt by introducing a claim but does not distinguish it from opposing views.</td>
<td>Effectively addresses the prompt by introducing a claim and distinguishing it from opposing views.</td>
<td>Effectively and insightfully addresses the prompt by introducing a precise claim and distinguishing it from opposing views.</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td>Demonstrates a lack of coherence, clarity, and cohesion.</td>
<td>Demonstrates limited coherence, clarity, and/or cohesion, making the writer’s progression of ideas somewhat unclear.</td>
<td>Demonstrates some coherence, clarity, and cohesion, and includes an introduction, conclusion, and logically grouped ideas.</td>
<td>Demonstrates a great deal of coherence, clarity and cohesion and includes an introduction, conclusion, and a logical progression of ideas.</td>
<td>Demonstrates purposeful coherence, clarity, and cohesion and includes a strong introduction, conclusion, and a logical, well-executed progression of ideas, making it easy to follow.</td>
</tr>
<tr>
<td>Item</td>
<td>0–2 points</td>
<td>3–4 points</td>
<td>5 points</td>
<td>6–7 points</td>
<td>8–9 points</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Claim Development/Use of Evidence</td>
<td>Provides little to no development of the claim; the development is not appropriate to the task, purpose, and audience.</td>
<td>Provides limited development of the claim by using some details, and/or text based evidence; the development is minimally appropriate to the task, purpose, and audience.</td>
<td>Provides development of the claim by using details, and/or text based evidence; the development is somewhat appropriate to the task, purpose, and audience.</td>
<td>Provides comprehensive development of the claim by using clear reasoning, details, and/or text-based evidence; the development is mostly appropriate to the task, purpose, and audience.</td>
<td>Provides effective and comprehensive development of the claim by using clear and convincing reasoning, details, and/or text based evidence; the development is consistently appropriate to the task, purpose, and audience.</td>
</tr>
<tr>
<td>Opposing Arguments</td>
<td>Does not address counter-arguments.</td>
<td>Demonstrates a limited understanding of the topic; does not adequately address counter-arguments.</td>
<td>Demonstrates an adequate understanding of the topic by presenting some counter-arguments.</td>
<td>Demonstrates a full understanding of the topic by presenting and refuting most counter-arguments with convincing reasoning.</td>
<td>Demonstrates a rich understanding of the topic by presenting and refuting all valid counter-arguments with convincing reasoning.</td>
</tr>
<tr>
<td>Language</td>
<td>Numerous errors in grammar and usage make the essay unclear. Style is mostly ineffective.</td>
<td>There are enough errors in grammar and usage to affect meaning. Inconsistently uses an effective style.</td>
<td>There may be errors in grammar and usage, but meaning remains somewhat clear throughout the essay. Uses a moderately effective style with some linking and transitional words, words to indicate tone, and/or domain-specific vocabulary.</td>
<td>There may be minor errors in grammar and usage, but meaning is mostly clear throughout the essay. Uses a mostly effective style with linking and transitional words, words to indicate tone, and/or domain-specific vocabulary.</td>
<td>There may be a few minor errors in grammar and usage, but meaning is clear throughout the essay. Uses an effective style with precise language, including linking and transitional words, words to indicate tone, and/or domain-specific vocabulary.</td>
</tr>
</tbody>
</table>
Section V. Suggestions for Refining Example Assessment #2 to Ensure Usefulness as a DDM

This promising DDM is designed to measure change over time in students’ ability to write coherent argumentative essays and in their ability to use text-based evidence to support a position. These writing assessments can be administered at key time points during the school year that are closely linked to intervals of instruction. They can be used by district educators to ensure a common performance assessment experience for students in all grade 9 English classes in that district.

Suggestions for refining this assessment are shared to illustrate some of the possible ways in which districts might, with slight modification, use existing assessments as DDMs.

**Suggestion 1: Collect evidence about the consistency of scoring.** This assessment is performance-based and is scored by trained educators who apply a rubric while considering each student’s written response. Districts using this approach will want to collect evidence that the prompt-based assessment can be scored reliably across teachers (inter-rater agreement or reliability) and over time (intra-rater reliability). To do so, they may want to consider collecting evidence about the degree to which (a) two educators scoring the same essay apply the rubric in a similar manner and assign the same score and (b) one educator applies the rubric consistently over time, assigning the same score to one student who submits writing samples at two closely spaced time points. They can use this information to improve training procedures and to put in place periodic checks on scorer calibration.

Districts may find it useful to collect student-developed essays that raters agree are representative of each level of performance for each criterion. These examples can be very helpful to educators when they are deciding if an essay should be scored at the low, mid or high point range according to the rubric. If they are strategically selected, and replaced when an essay emerges that is an even better match for a particular level of performance, use of representative exemplars can support educators in making the finely grained distinctions about student growth over time that are characteristic of a strong DDM. If they engage in such efforts, districts are likely to see that over time the criteria for low, moderate, and high growth will remain closely linked to the range of student responses that educators are most likely to see.

**Suggestion 2: Identify a second DDM that complements this promising DDM.** Since this assessment is performance-based, a district considering this approach may want to complement this measure by selecting a traditional assessment of knowledge and skills for use as a second DDM. For example, DDM developers might want to use results from the grade-appropriate Massachusetts Comprehensive Assessment System in writing and use the Student Growth Percentile estimates as their measure of growth.
Appendix A: Key ESE-Developed Resources to Support Districts with Implementing DDMs that Effectively Measure Student Growth

**August 2012**

*Part VII, Massachusetts Model System for Educator Evaluation: Rating Educator Impact on Student Learning Using District-Determined Measures of Student Learning*

Overview of DDMs and related concepts. It will be most valuable for districts beginning to learn about this work.

[http://www.doe.mass.edu/edeval/model/PartVII.pdf](http://www.doe.mass.edu/edeval/model/PartVII.pdf)

**Monthly Since February 2013**

*Educator Evaluator e-Newsletter*

Monthly newsletter designed to be a timely resource that provides key information, updates, and answers to frequently asked questions.

[http://www.doe.mass.edu/edeval/communications/newsletter/](http://www.doe.mass.edu/edeval/communications/newsletter/)

**March 2013**

*Introduction: District-Determined Measures and Assessment Literacy (Webinar Series #1)*

ESE has developed a nine-part webinar series on DDMs and assessment literacy. This series is targeted at district teams engaged in the work of identifying and selecting DDMs (e.g., district- and school-based curriculum and assessment leaders). Resources from these webinars include the recorded webinar and materials from each session.

[http://www.doe.mass.edu/edeval/ddm/webinar.html](http://www.doe.mass.edu/edeval/ddm/webinar.html)

**April 2013**

*Basics of Assessment and Assessment Options (Webinar Series #2 and #3)*

[http://www.doe.mass.edu/edeval/ddm/webinar.html](http://www.doe.mass.edu/edeval/ddm/webinar.html)

*Technical Guide A: Considerations Regarding District-Determined Measures*

Designed to increase assessment literacy by introducing foundational assessment concepts. It will be most valuable to districts interested in learning more about technical assessment concepts.

April 2013 (continued)

Assessment Quality Checklist and Tracking Tool

An interactive tool, built in Microsoft Excel, that organizes and catalogs information about individual assessments into a districtwide tracker of all potential DDMs. It will be most valuable to districts working to identify and select measures across the district.

http://www.doe.mass.edu/edeval/ddm/webinar/Quality-Tracking-Tool.xlsm

July 2013

Determining the Best Approach to District-Determined Measures (Webinar Series #4)

http://www.doe.mass.edu/edeval/ddm/webinar.html

DDM Technical Assistance and Networking Session I

ESE-hosted technical assistance and networking sessions intended to build on the Assessment Literacy Webinar Series and provide participants an opportunity to engage with colleagues from other districts around critical planning and implementation questions related to the piloting and eventual implementation DDMs.

http://www.doe.mass.edu/edeval/ddm/webinar.html

August 2013

Measuring Student Growth and Piloting District-Determined Measures (Webinar Series #5)

http://www.doe.mass.edu/edeval/ddm/webinar.html

September 2013

Technical Guide B: Measuring Student Growth & Piloting District-Determined Measures


DDM Technical Assistance and Networking Session II

http://www.doe.mass.edu/edeval/ddm/webinar.html

October 2013

Determining How to Integrate Assessments into Educator Evaluation: Developing Business Rules and Engaging Staff (Webinar Series #6)

http://www.doe.mass.edu/edeval/ddm/webinar.html

Using Current Assessments in District-Determined Measures: Leveraging the Curriculum-Embedded Performance Assessments from the Model Curriculum Units

http://www.doe.mass.edu/edeval/ddm/UsingAssessments.pdf
December 2013

Ramping Up for Next Year: Strategies for Using Current Assessments as DDMs (Webinar Series #7)
http://www.doe.mass.edu/edeval/ddm/webinar.html

DDM Technical Assistance and Networking Session III
http://www.doe.mass.edu/edeval/ddm/webinar.html

January 2014

Communicating Results (Webinar Series #8)
http://www.doe.mass.edu/edeval/ddm/webinar.html

February 2014

Sustainability (Webinar Series #9)
http://www.doe.mass.edu/edeval/ddm/webinar.html

Implementation Brief: Scoring and Setting Parameters
http://www.doe.mass.edu/edeval/ddm/Scoring-ParameterSet.pdf

Implementation Brief: Investigating Fairness
http://www.doe.mass.edu/edeval/ddm/Fairness.pdf

Implementation Brief: Using Student Growth Percentiles
http://www.doe.mass.edu/edeval/ddm/GrowthPercentiles.pdf

March 2014

Implementation Brief: Indirect Measures and Specialized Instructional Support Personnel (SISP)
http://www.doe.mass.edu/edeval/ddm/IMSISP.pdf

April 2014

Implementation Brief: Administrators
http://www.doe.mass.edu/edeval/ddm/Admin.pdf

Implementation Brief: Considerations for English Language Learners
http://www.doe.mass.edu/edeval/ddm/ELLEducators.pdf

Implementation Brief: Considerations for Special Education
http://www.doe.mass.edu/edeval/ddm/SpecialEduEducators.pdf
### Appendix B: Recommended External Resources on Measuring Student Growth: Guidance from the Research and Measurement Communities

Table 2. Research- and Measurement-Based Resources for Repeated Measures Approach

<table>
<thead>
<tr>
<th>Author</th>
<th>Resource</th>
<th>Topics Addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Katherine Castellano, University of California, Berkeley, and Andrew Ho, Harvard University, on behalf of CCSSO</td>
<td>A Practitioner’s Guide to Growth Models. (2012)</td>
<td>This guide provides cautions for districts electing to use a simple gain score approach (generalizing performance over multiple time points) (p. 19).</td>
</tr>
</tbody>
</table>