



**University of Vermont Department of Education
2020 Annual Report - Council for the Accreditation of Educator Preparation
Standard 4.1 – Impact on P-12 Student Learning and Development**

Executive Summary

The University of Vermont (UVM) Department of Education’s overall objective is to answer this primary question—**What impact do our graduates have on student learning growth?**

Sub-questions include:

- How do our graduates define impact on student learning?
- What types of teacher developed or standardized assessments do our graduates use to determine whether they have supported learning growth?
- What data do our graduates have available for analysis?
- How do our graduates modify instruction to meet learner needs based on these results?

The state of Vermont does not use value-added testing measures or link student results on state-wide tests to individual teachers, so we developed an alternate research plan that includes two components: 1) interviews with alumni (who are currently teaching in Vermont), and 2) evaluation of assessments and assessment data provided by the interviewees.

Part I on this report focuses on our academic year 2019-2020 study of 2017 graduates. Part II of this report presents results aggregated from three successive years of study: 2017-2018, 2018-2019, and 2019-2020.

Interviews, assessment samples, and sample student assessment data indicate that UVM graduates positively impact student learning growth. Most participants define student learning in terms of reaching proficiency on school-defined academic content and practices for learning targets anchored in state and national standards. Most also included development of social and emotional skills, habits of citizenship, or metacognitive skills.

Our graduates are using high-quality formative and summative assessments to gauge student academic knowledge, conceptual understanding, and skills in relation to learning targets, and are collecting data with a high potential to evaluate student learning growth. Based on participant perceptions of student learning growth in an academic unit and assessment data from that unit, we can state that our graduates are positively impacting student learning growth.

**University of Vermont Department of Education
2020 Annual Report - Council for the Accreditation of Educator Preparation
Standard 4.1 – Impact on P-12 Student Learning and Development**

Introduction

The University of Vermont (UVM) Department of Education’s overall objective is to answer this primary question—**What impact do our graduates have on student learning growth?**

Sub-questions include:

- How do our graduates define impact on student learning?
- What types of teacher developed or standardized assessments do our graduates use to determine whether they have supported learning growth?
- What data do our graduates have available for analysis?
- How do our graduates modify instruction to meet learner needs based on these results?

Given that the state of Vermont does not use value-added testing measures and we are not able to access student scores on state-wide tests tied to individual teachers, we developed an alternate research plan to examine the impact that our graduates have on student learning growth. This plan includes two components: 1) interviews with alumni (who are currently teaching in Vermont) focused on impact on student learning growth, and 2) collection and evaluation of assessments and assessment data provided by the interviewees.

Part I on this report focuses on our academic year 2019-2020 study of 2017 graduates. Part II of this report presents results aggregated from three years successive years of study: 2017-2018, 2018-2019, and 2019-2020.

In addition to the interviews and assessment data collection, the survey developed as part of our Selected Improvement Plan also addresses impact on student learning and there are open-ended questions in the Alumni Survey and the Employer Survey that address impact on student learning.

Part I: Academic Year 2019-2020 Study of 2017 Completers

Methods

Oversight of data collection and analysis related to the observation data and lesson plan analysis for Standard 4.1, as well as the iterative process of using results for continuous improvement is, provided by the CAEP Leadership Team composed of the Dean of the College, Associate Deans, Chair of the Department of Education, Director of Teacher Education, the College Director of Assessment, Data and Accreditation, and DOE Coordinator of Assessment and Accreditation.

Participant Selection

The sample for data collection in 2019-2020 for CAEP Standard 4.1 is from EPP graduates who completed their program in 2017 and are known to be employed in Vermont. At the time of data collection, most of these graduates were in their third year of teaching; new enough that their

practice would still reflect their formal professional preparation but no longer novices. Our goal was to create a sample of 10 or more alumni across all licensure programs in schools across the state. The sample was obtained by first narrowing the list of all 2017 EPP graduates (n=140) to those we were able to confirm were employed in a Vermont school in the fall of 2019.

Narrowing the list based on this criterion reduced the study population to 49 recent graduates. Beginning in October 2019, we invited all of the 2017 graduates in the study population to participate in this study. Invitations were sent by email—up to five times per individual—and included a brief description of the study. As incentive, participants were offered a \$25 Amazon gift card.

Ten 2017 graduates agreed to participate in the data collection for this study. Of these, eight were able to complete their study participation prior to implementation of COVID-19 related social distancing and school closures beginning in March 2020. (These same graduates also participated in the data collection for Standard 4.2.) The eight 2017 graduates who were able to participate in this study represent three of ten endorsement areas. Note that only six of ten endorsement areas are represented in the study population. The number of graduates from the study population in each endorsement area, the number in the sample, and the sample teaching grade levels are listed in Table 1 below. The schools in which study participants work represent four of Vermont's 14 counties. The distribution of schools by Vermont county is detailed in Table 2 below.

Table 1
Number of 2017 Completers by Endorsement Area
Study Population (N=49) vs. Sample (N=8)

Endorsement Area	Number of Graduates Confirmed as Teaching in VT	Number of Study Participants	Study Participants Teaching Grade(s)
Art	0		
Consulting Teacher	0		
Early Childhood	0		
Early Childhood Special Education	2		
Elementary Education	13	1	K
Middle Grades			
ELA	2	1	7 th – 8 th
Math	1		
Science	1		
Social Studies	4	2	7 th – 8 th
Music	1		
Physical Education	0		
Secondary Ed			
English	6		
Math	9	3	6 th – 9 th
Science	3	1	9 th – 12 th
Social Studies	4		
Special Education	3		

Table 2
Number of Study Participant Schools by Vermont County

County	Number Participants
Chittenden	4
Franklin	2
Washington	1
Windsor	1

Data Collection

All participants in the study shared plans for one lesson and were observed teaching that lesson (this was data collection for 4.2). Each participant was interviewed for 20 - 45 minutes about assessment and student growth for the instructional unit or grade level (as appropriate) of which the observed lesson was a part. The interviewer, Coordinator of Assessment and Accreditation for the University of Vermont's Department of Education, followed a semi-structured interview protocol which is included in the appendix. Interviews were audio-recorded and transcribed by the interviewer. Each of the participants also shared at least one example of an assessment from that unit and pre-post or other student assessment data related to the learning in the unit. Several also shared examples of student work on the assessments.

Data Analysis

Interview transcripts were coded using NVivo 12 software. Codes were developed from questions in the interview protocol, such as "Define Student Learning," "Assessment Tools," and "Data Usage." The data were further reduced through creation of display tables. Representative quotes were also selected.

Assessments and assessment data were analyzed using a rubric created for this study and included in the appendix. The rubric has three sections: Assessment Quality; Potential of the Assessment to Evaluate Impact on Student Learning; and Demonstration of Student Growth (do the data show growth in student learning?). Each section of the rubric will be described in more detail below with the findings in this report.

Findings

Interviews, assessment samples, and sample student assessment data indicate that 2017 UVM graduates positively impact student learning growth. What follows are findings about how 2017 graduates define student learning, the types of assessment tools they use, types of data available for analysis, how graduates use assessment data, and evidence that study participants are facilitating student learning growth.

What do graduates mean when they talk about student learning?

Vermont law requires that each supervisory union in the state establish proficiencies for middle and high school students to attain before graduation (<http://education.vermont.gov/student-learning/proficiency-based-learning>), and many PK programs and elementary schools are also in the process of adopting proficiency-based learning. Teachers who participated in this study are working in schools that have either fully implemented or are in the process of implementing proficiency-based systems.

Similar to previous years, most participants talked about student learning in terms of reaching proficiency on school-defined academic content and practices for learning targets anchored in state and national standards. Most also talked about student learning as an opportunity to engage with new content or experiences, and to extend or apply prior conceptions to new circumstances. They acknowledge that students are learning all the time, both in and out of school, and enter their classrooms with prior knowledge. Learning is the integration and application of existing and new ideas and experiences. As one teacher explained,

So, I really look at from where they started to where they ended, have they shown growth? Have they learned something from this unit that they can take to the outside world? Take to the next grade level? Take to another unit? Is there something they have comprehended and worked through that they can use in another setting, in another part of their life?

The idea of transfer to a different context was central to several of the definitions of student learning. One high school science teacher, for example, explained that student learning goes far beyond acquisition of academic content. She asserted that the three dimensions of the Next Generation Science Standards (<https://www.nextgenscience.org/>) can be applied to any discipline. Student learning in this sense is fundamentally about developing a systematic approach to inquiry and conceptual understanding. As she described it,

I think of student learning really as like can you do all of these things that are not just science, but practices and cross-cutting concepts really for any kind of thinking, no matter what field they want to go into.

Several participants also included dimensions of learning beyond academic content and practices. These dimensions were framed as social and metacognitive skills. One teacher summarized as follows:

So, student learning really means to me really influencing the whole person. So, helping them with their academics, but also, like how to use their words, how to take risks, how to be willing to try different things.

The dimensions of learning identified by participants and the frequency of the mention are detailed in Table 3 below.

Table 3
2020 Study of 2017 Completers
Participants' Descriptions of "Student Learning"

	What Does "Student Learning" Mean?	Number of Participants (out of 8) Who Mentioned This Dimension
Academic Content and Skills	Focus on Core Academic Content – Progress Toward Specific Targets Measured with Rubrics based on Standards	8
	Fluency in Discipline-Specific Practices	1
	Noticing Patterns and other Cross-cutting concepts.	2
	Class content sparks new inquiry into topics not necessarily covered in class.	1
	Transfer of new knowledge and skill to other contexts within and outside school. New learning integrated and remembered long after a test.	5
	Opportunity to grapple with rigorous content and learn through mistakes.	2
Social	Social and Emotional Skills	2
	Habits of Citizenship	1
	Building Trust and Connection	1
Meta cognitive	Feeling successful. Monitoring own growth.	2
	Executive Functioning	2

What types of assessments are graduates using?

Pre-Assessment

All study participants described using some form of pre-assessment. These include collection of baseline data on literacy skills in an early elementary class and baseline mathematical skills and understanding in a high school algebra class, as well as assessment of prior learning at the start of each new instructional unit. Five of the eight participants consistently use aligned, and in some cases identical, instruments for instructional unit pre-post testing. One high school math teacher is required to use a department common pre-post assessment for an algebra course. One middle level math teacher works in a school that is gradually switching from use of traditional textbook-based instructional materials to standards-based online materials. This teacher reports pre-post misalignment sometimes occurs due to decisions at the department level to use unit pre-tests from the textbook and post-tests based in the online materials and is keenly aware that "our assessments are a bit funky right now in this transition." Most pre-assessments are paper-pencil activities, but one high school science teacher reports engaging her students in trivia games to pre-assess content knowledge. She also uses graphing or other relevant activities to pre-assess proficiency with specific science practices and cross-cutting concepts.

Ongoing Formative Assessment

Formative assessments used throughout instructional units include entrance tickets and exit tickets, other formal progress monitoring such as quizzes or non-summative assignments, and

informal observation and monitoring of class work. Entrance tickets are used primarily to activate student recall of prior learning, in part through reflection on what is working well so far to support their learning and suggestions for change. Exit tickets are used primarily to gauge student understanding after a lesson or series of learning opportunities, and results are usually assessed in relation to a proficiency scale and recorded for further analysis. At the kindergarten level, formal progress monitoring includes literacy assessments related to programs like *Jolly Phonics*, and at the middle and high school levels takes the form of quizzes and weekly assignments. All participants described using informal or on-the-spot formative assessment through observation of student work during class, homework check-ins, and/or asking students to self-assess their current level of confidence or understanding by showing thumbs up or down. Such informal assessment is ongoing throughout instructional time. As one teacher described it, “There’s always that informal check-in that’s giving you informal feedback you’re kind of holding in your head.”

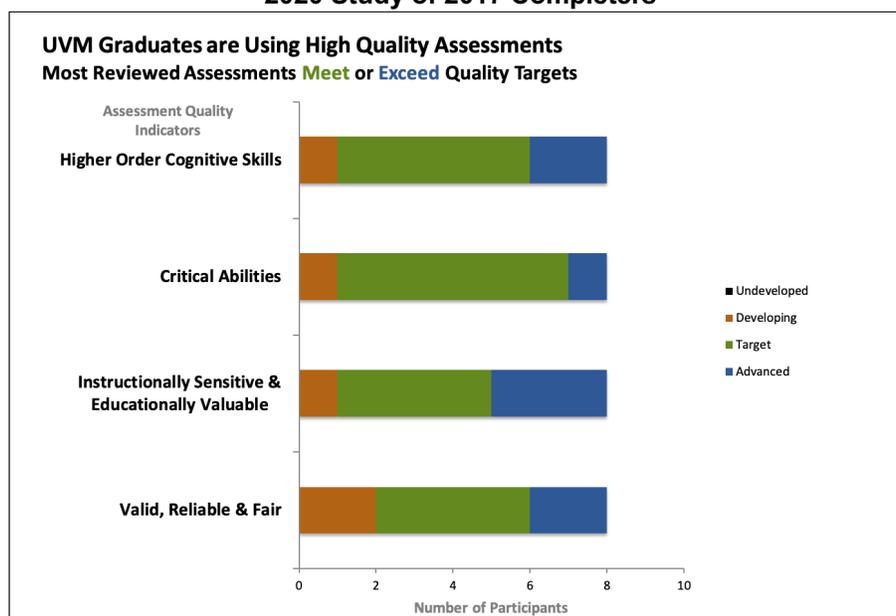
Summative Assessment

Similar to previous reporting years, summative assessments can still take the form of “traditional” multiple-choice, short answer, and essay exams, and most participants this year reported using summative performance tasks as well. In one middle level math class, for example, in a unit on proportions students were assessed both with a paper and pencil assessment and by their ability to build tiny homes that demonstrated understanding of scale, area, and surface area. A different middle level math teacher reported periodically foregoing the traditional unit assessment and replacing it with an end-of-unit project. In a recent example, students were asked to make a scale model of any object they wanted to work with, and then write about their process: the original measurements; their process for scaling; and the scaled measurements. Several participants mentioned that developing creative summative assessments, potentially co-designed with students, is an area for professional growth. In proficiency-based systems students are generally allowed to re-take summative assessments until they can demonstrate proficiency. Several participants noted that they do require students to meet with them to engage in additional learning before allowing re-takes. As noted earlier in this section, most summative assessments are aligned with unit or course pre-assessments.

What is the quality of assessments shared by participant for this study?

As described above, participants shared at least one example of an assessment they used in their then current or recent instructional unit. These samples were used to evaluate the quality of assessments used by recent graduates. As indicated in Chart 1 below, the assessments used by study participants are overall high quality. All eight of the sample assessments met or exceeded the target on at least two of the four dimensions, and six of the sample assessments met or exceeded the target on all four dimensions.

Chart 1
2020 Study of 2017 Completers



The rubric used to evaluate assessment quality was based on the Stanford Center for Opportunity Policy in Education's *Criteria for High Quality Assessment* (2013), which draws heavily from the work of Norm Webb (2002 & 2005). The rubric is included in the appendices of this report. The four dimensions of the rubric are:

- Assessment of Higher-Order Cognitive Skills
- High-Fidelity Assessment of Critical Abilities – Research, Analysis & Synthesis; Experimentation & Evaluation; Communication & Collaboration; and Complex Problem-Solving
- Use of Items that are Instructionally Sensitive and Educationally Valuable
- Assessments are Valid, Reliable, and Fair

What types of data do graduates have for analysis?

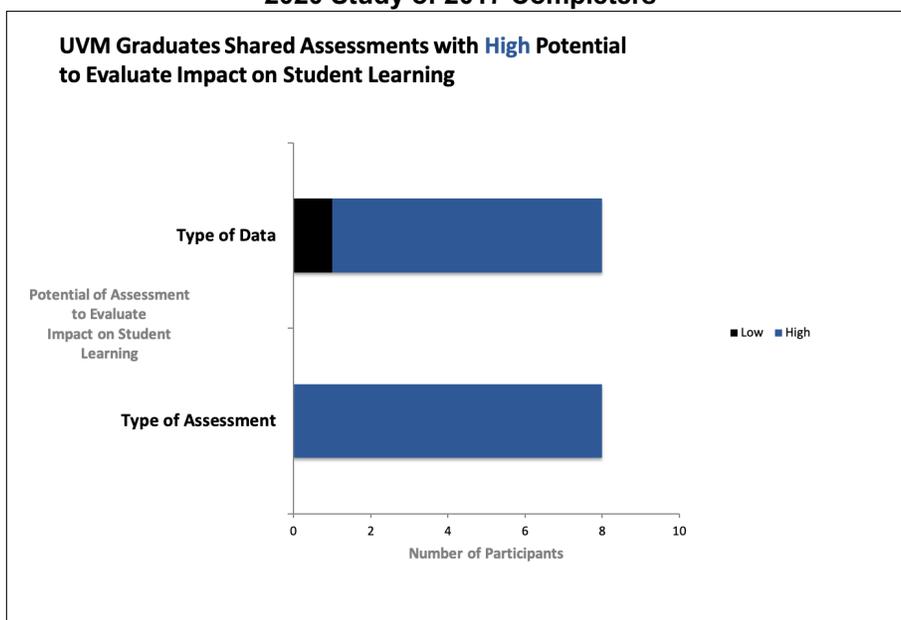
Participants gathered both qualitative and quantitative data from the assessments described above. Qualitative data include observations of student responses and reactions during in-class questioning, discussion, activities, and interactions. Participants report little or no formal tracking of this type of data, although one high school teacher essentially records her observations in the school's online system for scheduling student call-backs for additional academic support. Quantitative data include scores on traditional multiple-choice or other paper and pencil tests, and numbers or percentages of students who are below, at, or above proficient on particular learning targets. While two participating teachers record proficiency levels as B, D, P, or E (Below, Developing, Proficient, or Expanding), in most proficiency-based systems teachers essentially quantitize qualitative data when they rate proficiency on a scale of one to four (one-below, two-nearly, three-proficient, four-exceeding). These differences are likely rooted in school-level approaches to proficiency-based grading and the constraints of online

student data management systems. At least one participating teacher also tracks whether proficiency on a summative assessment was demonstrated on the initial or follow-up attempt.

As noted in past reporting cycles, except for assessments associated with commercial elementary literacy programs, most participants reported using detailed rubrics or scales to rate each student’s proficiency level on each learning target connected to some formative and most summative assessments. Usually these scales have been collaboratively developed with a combination of team or department colleagues, instructional coaches, and/or a curriculum coordinator. Several participants work in schools that use online proficiency-based grading systems which link assignments and assessments to specific learning targets, making it easier to see how often each target is assessed, as well as tracking student progress over time. Several participants this year also keep their own spreadsheets to track student progress across a unit, and two use hand-drawn grids on sheets of paper. Whether using an online system or simply tracking in a spreadsheet, participants report using the data to monitor student progress toward proficiency. Sometimes they analyze data individually, but often as part of school-based teaching or data teams.

The sample assessments and assessment data shared by study participants were rated on their potential to evaluate impact on student growth. All eight of the sample assessments – performance tasks, unit tests, kindergarten literacy skill acquisition tracking, unit post-tests/summatives – were deemed to have high potential. Most of the data types – generally class (by de-identified student) proficiency scores over the course of a unit or unit pre-post – were also deemed to have high potential. The teacher whose data were deemed to have low potential used inconsistent scoring methods across the unit (points on homework but proficiency scales on exit tickets) and explained that exit ticket questions correspond directly to increasingly higher levels on the learning target proficiency scale, so apparent growth across time for some students might actually be the result of ceiling effects on the early exit tickets.

Chart 2
2020 Study of 2017 Completers



What do graduates learn from the data and how do they use it?

Similar to previous years of this study, and consistent with their definitions of student learning, participants report that the data they collect help them to build understanding of where their students are in relation to learning targets and to consider how best to facilitate continual growth. They use observational data for in-the-moment decisions about, for example, follow-up questions, further discussion, clarifications, encouragement, or explicit discussion of transferable skills. Formative assessments not only gauge student learning across a class, but are also an opportunity to give individualized feedback, usually with explicit links to scoring rubrics. One middle level teacher summarized what she learns from assessment data this way:

Where the students are. If they are proficient, if they are developing, if they are beginning. Also, it gives me an idea if it's an entire class that is not proficient on that topic, if I need to reteach anything or everything at the same time, depending on where the class is at, what the general overall understanding is. But it is also letting me know how that specific student learns and if that specific student continues to be beginning, we probably need to put support in place to help that student become proficient.

Data inform which instructional resources participants use and the types of choices they make available to students, including centers for play-based learning at the kindergarten and recommendations for supplemental supports at the middle and high school levels. Data are also the basis for communication with students and parents about student progress.

Several participants describe using a more-or-less formal backward design (Wiggins & McTighe, 2005) process for unit planning. Learning targets for a unit or span of learning are clearly identified, and the summative assessments are, if not finalized, at least outlined prior to engaging students in learning opportunities. Rather than follow a rigid path toward a learning target, however, participants use formative data to support ongoing instructional decision-making. One middle level teacher explained that she has learned to build flexibility into unit plans so she can respond to what she sees in formative assessment results.

I like planning ahead, but I had a few days that just said in my plan, mini-lesson on whatever. So, we were able to put in - the literacy coach especially was concerned with the conclusion thing, so she said we were going to put in an extra day on this. So, we were able to add in pieces to our instruction, which I didn't do last year because I was so hyper-focused on having a plan for every single day that I was really afraid to get off that schedule.

This teacher, along with several others, also talked about using assessment results to check for retention of prior learning. One high school teacher periodically includes topics from previous units on exit tickets to check for retention.

Sometimes on an exit ticket I'll put a question that we did in the previous unit, or that we've done a while ago, and I'll just try to see, are we retaining that information as a group? Is it something we need to come back to? And I had put one on today that was similar to one we had just done, and they called me out on it. "We already know how to do this!"

This teacher is implementing a form of “interleaving” (Brown, Roediger, & McDaniel, 2014) through her formative assessments; as she checks for long-term skill and understanding, she is also helping students to build long-term retention by providing periodic opportunity to revisit content. Similarly, a high school science teacher uses assessment to systematically revisit previous learning to plan for instruction that provides students multiple opportunities to engage with context, practices, or cross-cutting concepts.

We know CCC 4 specifically – systems and system models – I know that the majority of students are not proficient on it, and those that are proficient have only shown proficient once on it. But a lot of them are not understanding it. So, now I'm wrapping that into my next unit – the chemistry unit that we just started it's also going to have CCC 4. Which is wasn't originally going to.

Most participants are also using assessment to empower students to track their own learning. Strategies for this include reflection questions on entrance or exit tickets, exit tickets that indicate the level or proficiency represented by a correct answer and where a student is expected to be at that point in a unit, and pair-decks that allow students to anonymously compare and analyze the work of other students in a class. One teacher meets with students at the end of each unit to compare pre- and post-assessment results to help students see their progress.

When asked for advice to new teachers who are trying to assess their own impact on student learning growth, most participants again this year talked about the value of regularly collecting, recording, and using a variety of data to tailor learning opportunities for their students. Several participants emphasized the importance of the initial data point: pre-assessment or baseline data is necessary for measuring growth. The variety of data should include whether students are happy and engaged, and their opinions about what kinds learning and assessment opportunities work well for them (and which do not). Participants would remind new teachers that learning can take a long time, and is enhanced through positive teacher-student relationships. One participant this year stressed that for assessment results to be useful, students must be allowed to grapple with problems independently. Another suggested impact can be measured by the extent to which students have opportunities to express themselves through choices and be creative in what they are doing with the content, as opposed to the teacher directing everything. Another would remind new teachers that not all learning is formally assessed, and to notice social and metacognitive skill development. This participant also suggested that new teachers avoid taking either too much or not enough credit for student learning growth.

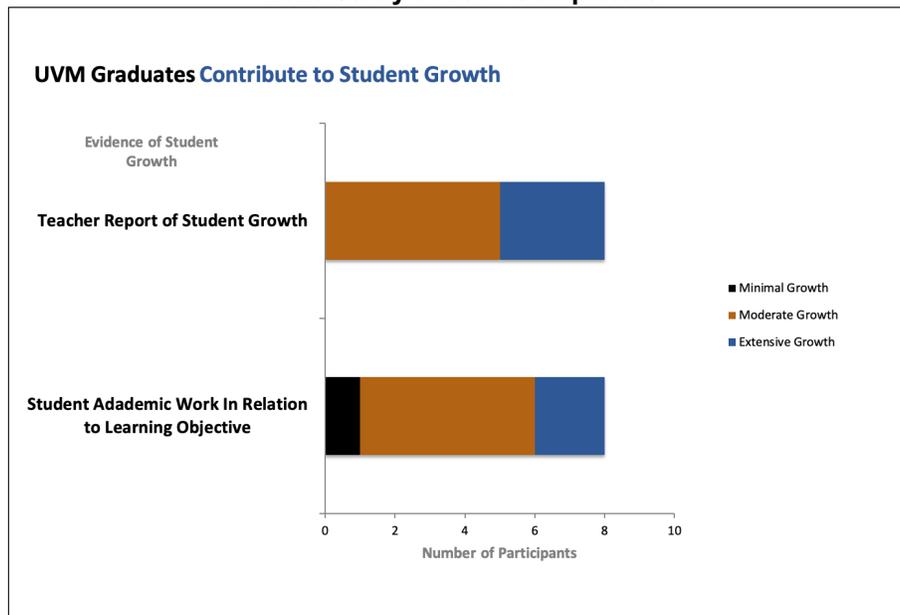
What impact do graduates have on student learning growth?

The combination of interview and sample assessment data indicates that study participant contribution to student learning growth is moderate to extensive. To make this determination, interview and assessment data were rated on the following scale:

- Minimal Growth
 - Teacher reports little or no learning
 - Student work far below the standard or learning objective
- Moderate Growth
 - Teacher reports moderate learning
 - Student work is approaching the standard or learning objective
- Extensive Growth
 - Teacher reports extensive learning
 - Student work meets or exceeds the standard or learning objective

In one case, teacher reporting during the interview matches student data which show growth over a period of time on academic learning targets and that students met or exceeded the learning target. In three cases, teacher reporting during the interview matches student data which show moderate growth and student learning approaching the learning target. In one case the teacher reported moderate growth but later shared data showing extensive growth by the end of that unit. In two cases the teacher reported extensive growth but about half the student work was below target levels. In one case, a course for students who struggle with the math, the teacher reported moderate growth but data show most student work well below the learning target.

Chart 3
2020 Study of 2017 Completers



Conclusion

UVM EPP 2017 graduates in this study sample are positively impacting student learning growth. Data available mid-way through the 2019-2020 academic year from teacher interviews and overall high-quality assessments show most students demonstrating moderate to extensive growth toward learning targets. It is clear from the interviews (and classroom observations) that formative assessment is fully integrated into the daily routines of the study participants, and that data collection and analysis are growing dimensions of the school cultures. As proficiency-based learning takes hold in Vermont, most of the study participants report working on instructional teams that collaborate to refine learning targets and assessment scales, review student data, and develop instructional responses to enhance student learning. Several participants identified development of performance assessments as a growth area. Several shared that their UVM content-focused methods courses deeply impacted their current classroom practice, and others noted the importance of their preparation to collaborate with school colleagues. Study participants expressed appreciation for the foundation in assessment for teaching and learning embedded in their UVM EPP experiences, and suggest that future graduates might benefit from even more exposure to strategies for data collection, analysis, and use in proficiency-based learning systems

Part II: Results from Three Cycles of Data Collection

Methods

The same participant selection, data collection and data analysis methods were used in each of the three academic years (AY 2017-2018, AY2018-2019, and AY 2019-2020) we conducted this study, and the same rater conducted the interviews and assessment collection each year. Those methods are described in detail in Part I one of this report.

In each of the data collection cycles we met our study sample size goal of ten program completers. In the first year of the study we had eleven participants, in the second year ten, and in the third year ten who agreed to participate – with eight able to complete the study prior to school closures due to the COVID-19 pandemic. These participants represent seven of the ten endorsement areas. The aggregate number of graduates in the three study populations in each endorsement area, the aggregate number in the sample, and the aggregate sample teaching grade levels are listed in Table 4 below. Schools in which study participants work represent nine of Vermont's 14 counties. Distribution of schools by Vermont county is detailed in Table 5 below.

Table 4
Three-Year Aggregated (2015, 2016, & 2017) Completers by Endorsement Area
Study Population (N=127) vs. Sample (N=29)

Endorsement Area	Number of Graduates Confirmed as Teaching in VT	Number of Study Participants	Study Participants Teaching Grade(s)
Art	3		
Consulting Teacher	4		
Early Childhood	5		
Early Childhood Special Education	10	2	PK
Elementary Education	27	5	PK-6 th SPED; K; 1 st ; 4 th -5 th
Middle Grades			
ELA	3	2	7 th – 8 th
Math	3	1	6 th
Science	2	1	4 th
Social Studies	4	2	7 th – 8 th
Music	3	1	K-6 th
Physical Education	1	1	K-6 th
Secondary Ed			
English	12	1	9 th -12 th
Math	14	6	9 th -12 th
Science	12	5	7 th -12 th
Social Studies	11	1	9 th -12 th
Special Education	13	1	K-8 th

Table 4
Number of Study Participant Schools by Vermont County

County	Number Participants
Addison	1
Caledonia	1
Chittenden	11
Franklin	9
Grand Isle	1
Lamoille	1
Rutland	1
Washington	2
Windsor	2

Findings

Interviews, assessment samples, and sample student assessment data indicate that 2015, 2016, and 2017 UVM graduates positively impact student learning growth. What follows are brief summaries of findings about these graduates define student learning, the types of assessment tools they use, types of data available for analysis, how graduates use assessment data, and evidence that study participants are facilitating student learning growth.

What do graduates mean when they talk about student learning?

Vermont law requires that each supervisory union in the state establish proficiencies for middle and high school students to attain before graduation (<http://education.vermont.gov/student-learning/proficiency-based-learning>), and many pre-K programs and elementary schools are also in the process of adopting proficiency-based learning. Most teachers who participated in our annual studies work in schools that have either fully implemented or are in the process of implementing proficiency-based systems.

Likely because they work in proficiency-based systems, across all three years most participants talked about student learning in terms of reaching proficiency on school-defined academic content and practices for learning targets anchored in state and national standards. Most also talked about student learning as an opportunity to engage with new content or experiences and extend or apply prior conceptions to new circumstances. They acknowledge that students are learning all the time, both in and out of school, and enter their classrooms with prior knowledge. Learning is the integration and application of existing and new ideas and experiences.

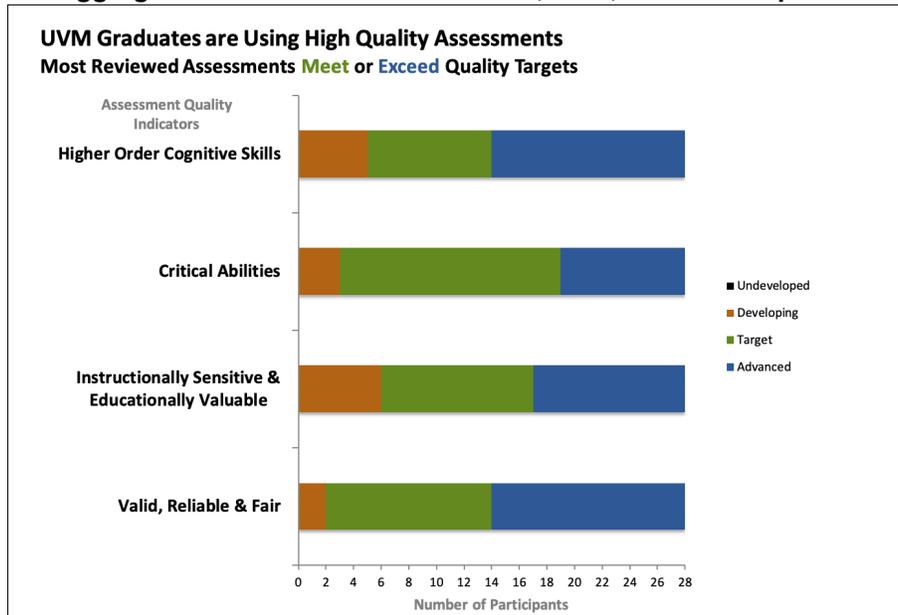
Most participants also included dimensions of learning beyond academic content and practices. These dimensions were framed as social and emotional skills, habits of citizenship, or metacognitive skills.

What types of assessments are graduates using?

Across all three years, participants describe using formative and summative assessments to help them gauge student academic knowledge, conceptual understanding, and skills in relation to learning targets. Formative assessment includes formal pre-assessment prior to an instructional unit, as well as ongoing checks-ins in the form of entrance and exit tickets, small group or whole class discussion, and observation of students as they engage in other learning experiences. Summative assessments include both traditional pencil-paper unit tests and projects or other non-traditional demonstrations of proficiency on specific standards.

Each participant shared at least one example of an assessment they used in their then current or recent instructional unit. These samples were used to evaluate the quality of assessments used by recent graduates. As indicated in Chart 4 below, the assessments used by study participants are overall high quality. All of the sample assessments met or exceeded the target on at least two of the four dimensions, and 17 of the sample assessments met or exceed the target on all four dimensions.

Chart 4
Aggregated Data from Studies of 2015, 2016, & 2017 Completers



The rubric used to evaluate assessment quality was based on the Stanford Center for Opportunity Policy in Education's *Criteria for High Quality Assessment* (2013), which draws heavily from the work of Norm Webb (2002 & 2005). The rubric is included in the appendices of this report. The four dimensions of the rubric are:

- Assessment of Higher-Order Cognitive Skills
- High-Fidelity Assessment of Critical Abilities – Research, Analysis & Synthesis; Experimentation & Evaluation; Communication & Collaboration; and Complex Problem-Solving
- Use of Items that are Instructionally Sensitive and Educationally Valuable
- Assessments are Valid, Reliable, and Fair

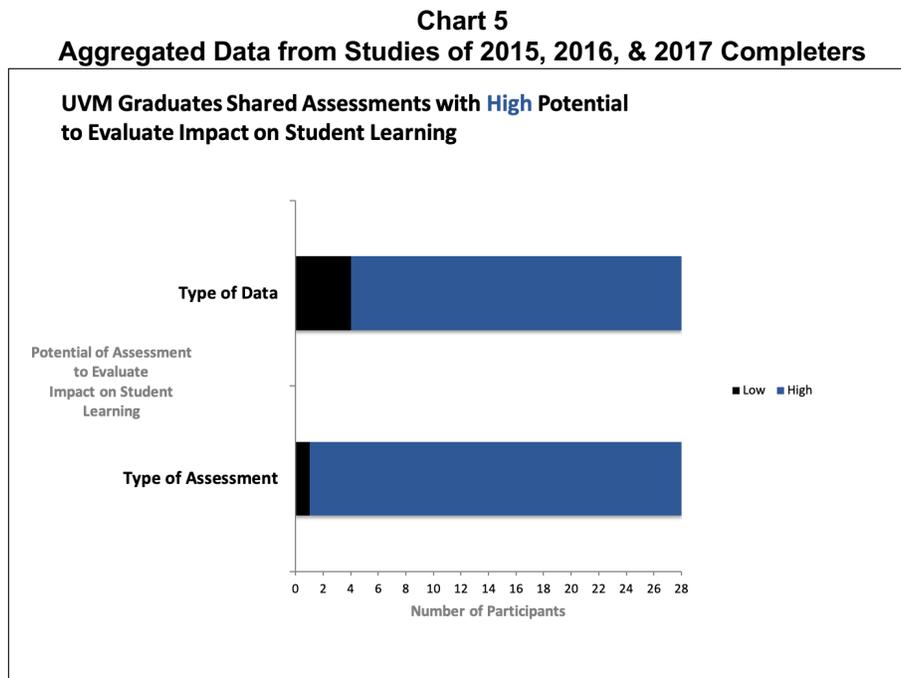
What types of data do graduates have for analysis?

Participants gathered both qualitative and quantitative data from the assessments described above. Qualitative data include observation of student responses and reactions during in-class questioning, discussion, activities, and interactions. In some cases, participants reported making notes of these observations in a formal electronic reporting system or other format, but this type of data is mostly unrecorded. Quantitative data include scores on formal formative and summative assessments, including entrance and exit tickets, traditional multiple-choice tests, and other forms of summative assessments. Increasingly over the three years, scores were recorded as numbers or letters on proficiency scales.

Except for assessments associated with commercial elementary literacy programs, most participants reported using detailed rubrics or scales to rate each student's proficiency level on each learning target connected to some formative and most summative assessments. Usually

these scales have been collaboratively developed with a combination of team or department colleagues, instructional coaches, and/or a curriculum coordinator. Several participants work in schools that use online proficiency-based grading systems (such as *JumpRope*, *Infinite Campus*, or *Power School*) which link assignments and assessments to specific learning targets, making it easier to see how often each target is assessed, as well as tracking student progress over time. Whether using an online system or simply tracking in a spreadsheet, participants report using the data to monitor student progress toward proficiency. Sometimes they do this individually, but often as part of school-based teaching or data teams.

As illustrated in Chart 5 below, sample assessments and assessment data shared by study participants were rated on their potential to evaluate impact on student growth. Twenty-seven of the sample were deemed to have high potential. Most of the data types – generally class (by de-identified student) proficiency scores over the course of a unit or unit pre-post – were also deemed to have high potential.



What do graduates learn from the data and how do they use it?

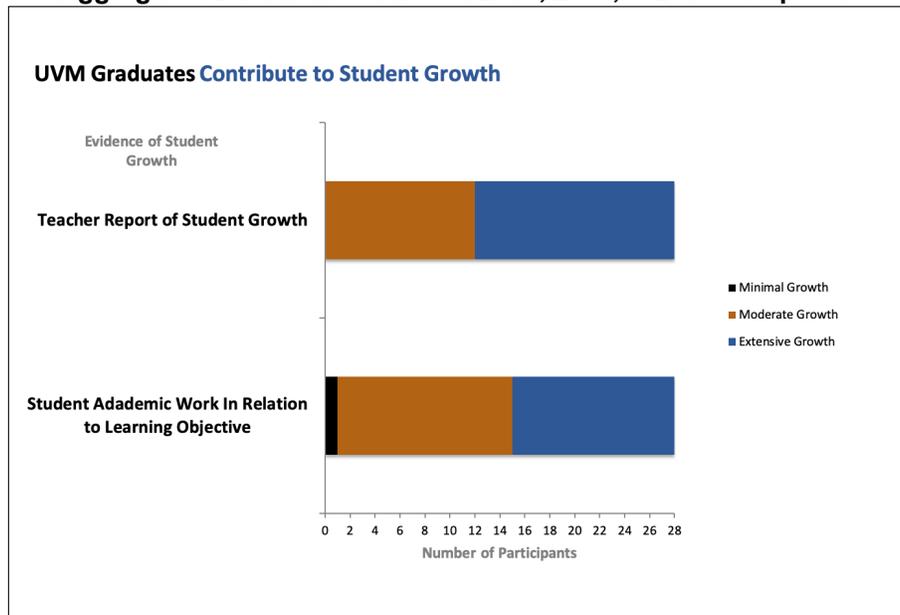
Across all three study years, and consistent with their definitions of student learning, participants report that the data they collect help them to build understanding of where their students are in relation to learning targets and consider how best to facilitate continual growth. This is the case both for classes as a whole, and for individual students. Most study participants stressed the importance of positive teacher-student relationships, and emphasized that really knowing students allows a teacher to use data to tailor instruction to meet students where they are, as opposed to where they are expected to be on a learning progression.

What impact do graduates have on student learning growth?

The combination of interview and sample assessment data indicates that study participants are contributing to moderate to extensive student learning growth (Chart 3). To make this determination, interview and assessment data were rated on the following scale:

- Minimal Growth
 - Teacher reports little or no learning
 - Student work far below the standard or learning objective
- Moderate Growth
 - Teacher reports moderate learning
 - Student work is approaching the standard or learning objective
- Extensive Growth
 - Teacher reports extensive learning
 - Student work meets or exceeds the standard or learning objective

Chart 6
Aggregated Data from Studies of 2015, 2016, & 2017 Completers



In most cases, teacher verbal reflection on student growth matched the student data they provided. In ten cases, teacher reporting during the interview matches student data which show that students met or exceeded the learning target. In 9 cases, teacher reporting during the interview matches student data which show moderate growth and student learning approaching the learning target.

Conclusion

Across all three years of study we find that UVM's EPP graduates are positively impacting student learning growth. Teacher interviews and results of the overall high-quality assessments they are using show most students demonstrating moderate to extensive growth toward learning targets. By midway through their third year in the field, most now have enough experience in the

classroom to deeply understand the critical connection between assessment and instruction. It is clear from the interviews (and classroom observations) that formative assessment is fully integrated into the daily routines of most study participants, and that data collection and analysis are growing dimensions of the school cultures. As proficiency-based learning takes hold in Vermont, most of the study participants report working on instructional teams that collaborate to refine learning targets and assessment scales, review student data, and develop instructional responses to enhance student learning.

References

- Brown, P., Roediger, H., & McDaniel, M. (2014). *Make it stick: The science of successful learning*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Darling-Hammond, L., Herman, J., Pellegrino, J., et al. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Wiggins, G. & McTighe, J. (2005). *Understanding by Design*. Alexandria, VA: ASCD Publishers.

Appendices

UVM CEAP 4.1 Interview Protocol

Script – Overview and Purpose of the Interview

Thank you for agreeing to participate in this interview!

As stated in the Information Sheet, the purpose of this interview is to collect information on how graduates of UVM’s teacher education programs impact student-learning growth. The information will be used as part of a continuous improvement process for all of our teacher preparation programs and as evidence in reporting to our accrediting body the Council for the Accreditation of Educator Preparation. Your responses will be compiled with those of other interviewees and will not be personally identifiable in any sharing of results. Our conversation will be recorded and transcribed. All data will be kept on password protected devices computers.

Please do not use any real names, including the name of your school and school system, your students, and your colleagues, when responding to the prompts. If you inadvertently use a name, the name will be omitted from the transcription.

We hope that participation in this interview will be an opportunity to reflect on your professional practice as a K-12 teacher.

Please share with us assessments you described in the unit example and the resulting student data you collected and anything else you think would be helpful to us in understanding how you used assessment to improve student learning. You can either share paper copies of material now or send files via email.

	Semi-Structured Focus Group Questions
Intro & Defining “ Student Learning”	<p>Introduction</p> <ul style="list-style-type: none"> • <i>Overview of the Purpose of the focus group (See Script)</i> <p>In my experience talking with teachers, I know that “student learning” can be defined in many different ways. How do you define student learning?</p>
Assessment in one instructional unit	<p>Think about one specific instructional unit you taught this year. It can be in any subject area. If I had been observing in your classroom during that unit, what would I have seen you doing to assess student learning?</p> <p><i>Prompts</i></p> <ul style="list-style-type: none"> • <i>Tools, instruments, processes (standardized or self made)</i> • <i>Pre-assessments, On-the-spot formative assessment, Student discussion, Full class question – answer sessions, Homework review, Students presenting their work to the class, Quizzes, Unit tests</i> • <i>Can you tell me more about that?</i>

<p>Data Analysis and Use</p>	<p>Thinking again about that same instructional unit, how did you <u>analyze</u> the data on student learning that you collected? Please give as least two examples.</p> <p><i>Prompts</i></p> <ul style="list-style-type: none"> • <i>Can you tell me more about that?</i> <p>What kinds of information did these tools and instruments give you about student learning?</p> <p><i>Prompts</i></p> <ul style="list-style-type: none"> • <i>Can you tell me more about that?</i> <p>How did you <u>use</u> the data that you collected on student learning?</p> <p><i>Prompts</i></p> <ul style="list-style-type: none"> • <i>Plan instruction? Modify instruction? Student grouping? Summative grading?</i> • <i>Can you tell me more about that?</i> <p>How would you summarize your students' learning in that unit? Please give examples.</p> <p><i>Prompts</i></p> <ul style="list-style-type: none"> • <i>Examples of content learning?</i> • <i>Examples of growth in content specific practices?</i> • <i>Examples of growth in transferable skills?</i> • <i>Can you tell me more about that?</i>
<p>Closing</p>	<p>If you were giving advice to new teachers about how to assess their own impact on student learning, what would you say?</p> <p><i>Prompts</i></p> <ul style="list-style-type: none"> • <i>What have you learned about teaching and assessment that you think is really important to share with others?</i> • <i>How do <u>you</u> know if your instruction is helping students?</i> <p>Is there anything we did not ask about that you think is important as we build a picture of how UVM graduates understand their impact of student learning?</p> <p>Do you have any questions for us at this point?</p> <p>Thank you!</p>

UVM CAEP 4.1 - Rubric to evaluate participant assessments and assessment data

Part A: Assessment Quality

Based on Stanford Center for Opportunity Policy in Education's *Criteria for High Quality Assessment (2013)*, which draws heavily from the work of Norm Webb (2002 & 2005).

Performance Indicator	Area of Concern (1)	Approaches Standard (2)	Meets Standard (3)	Exceeds Standard (4)	Rating (1-4)
Assessment of Higher-Order Cognitive Skills (Using Webb's (2002) Depth of Knowledge Levels) (HQA Standard 1)	The assessment consists primarily of Level 1 (list, identify, calculate, etc.) tasks.	The assessment goes beyond Level 1 tasks and includes some Level 2 (organize, estimate, classify, etc.) and/or Level 3 (explain, revise, construct) tasks.	No more than one-third of the items are Level 1 tasks and at least half of the assessment is comprised of Level 3 and Level 4 (critique, design, prove) tasks.	The assessment allows for extensive demonstration of conceptual knowledge and transferable skills with most items at Levels 3 & 4.	
	Comments:				
High-Fidelity Assessment of Critical Abilities – Research, Analysis & Synthesis, Experimentation & Evaluation, Communication, Collaboration, & Complex Problem-Solving (Using Webb's Alignment Tool, 2005) (HQA Standard 2)	The assessment consists primarily of retelling narrative, mathematical computation, labeling, or other routine procedures. Questions/tasks are primarily selected response items.	The assessment consists primarily of selected response items that ask students to summarize text, formulate and/or solve routine multi-step problems, identify patterns, describe cause and effect, or organize, represent, and interpret data.	The assessment includes some authentic tasks in which students must support ideas with examples and details, use appropriate voice, craft research questions and design investigations, develop a model for a complex situation, determine author purpose and implications for interpretation, or apply a concept to a different context.	The assessment consists primarily of authentic tasks or projects that includes: problem identification, experimental design, data analysis and reporting of results; design and/or application of mathematical model to a situation; analyze & synthesize information from multiple sources.	
	Comments:				
Use of items that are Instructionally Sensitive and Educationally Valuable (HQA Standard 4)	The assessment primarily measures skills or content that students are likely to have learned outside the context of the instructional unit, and/or test-taking skills. The purpose of the assessment is unclear and the results are not useful for guiding instruction.	The assessment measures some instructional content, but also measures outside experiences and learning and is therefore only somewhat useful to guide instruction.	The assessment reflects intended learning outcomes from the unit, and success does not depend on learning or experiences in other contexts. The purpose of the assessment is to measure growth and results are useful for guiding instruction.	The assessment reflects intended learning outcomes, success does not depend on learning in other contexts, and completing the task(s) extends student learning. The purposes of the assessment are to measure growth and extend learning, and results guide future instruction.	
	Comments:				

Performance Indicator	Area of Concern (1)	Approaches Standard (2)	Meets Standard (3)	Exceeds Standard (4)	Rating (1-4)
Assessments are Valid, Reliable and Fair. (HQA Standard 5)	Assessment does not match skills and content of the unit/lesson/standard. The questions/tasks contain bias and may not accurately represent student understanding and skill.	Assessment partially matches skills and content of the unit/lesson/standard, and/or the questions/tasks contain bias and may not fully represent student understanding and skill.	Assessment matches skills and content of the unit/lesson/standard. The questions/tasks are unbiased and are likely to accurately represent student understanding and skill.	Assessment matches skills and content of the unit/lesson/standard. The questions/tasks are unbiased. Results likely to accurately represent student understanding and skill and predict real-world success in the domain.	
	Comments:				

Part B: Potential of the Assessment to Evaluate Impact on Student Learning

Indicator	Low Potential	High Potential	Rating
Type of Assessment	Pre-test only On-the-spot formative assessment	Performance Task Unit assessment Portfolio	
	Comments:		
Type of Assessment Data	Global Score Percent Correct	Pre-Post Item Level Item Level data Teacher evaluation of student work	
	Comments:		

Part C: Demonstration of Student Growth (do the data show growth in student learning?)

Indicator	Minimal Growth (1)	Moderate Growth (2)	Extensive Growth (3)	Rating (1-3)
Teacher report of student growth (from interview.)	Teacher reports little or no learning.	Teacher reports moderate learning.	Teacher reports extensive learning.	
	Comments:			
Student work in relation to the learning objective.	Student work is far below the standard or learning objective.	Student work is approaching the standard or learning objective.	Student work meets or exceeds the standard or learning objective.	
	Comments:			