

South Seneca Central School

Assessment Audit



Teaching is the Core Committee

Patricia Parsons
Cindy Brewer
Stacey Clark
Adam Rundell
Kathy Sinkiewicz
Cindy Woodard

2015

Table of Contents

Audit Report

I.Rationale.....2

II. Guiding Questions.....3

III. Methodology.....3

IV. Data Analysis.....5

V. Recommendations.....13

VI. Future Plans.....14

VII. Conclusions.....15

Appendix

Summary of Assessment Audit Results.....16

I. Rationale.

Our main entry point for the 2014-2015 assessment audit was to determine the Impact on Instruction of our assessments. The Impact on Instruction entry point is broken down into three categories: feedback, instructional decisions, and diverse learners. Our group was interested in this because we really felt that as a whole, our school genuinely puts our kids and their well-being above all else. Because of this, we want to be sure that our assessments are benefiting our children's instruction. Other entry points that we identified as important and noteworthy were: Alignment and Validity, Diversified and Balanced, and Reliability.

Team members requested any and all assessments that colleagues were willing to give for analysis. A brief explanation of the Teaching is the Core grant was given, but we didn't request certain assessments in fear that it might skew our results. Although we love to see the amazing work that our colleagues do, we wanted to see real, everyday assessments that were being given in our children's classrooms. Also, we did not ask only for summative assessments, but requested any type of assessment that may be given in class. Team members took an inventory of submitted assessments, review the components of the assessments, and explored patterns across the many assessments from different grade levels and different subject areas. The audit team used a rubric to evaluate how each assessment is linked to the entry points identified by the grant.

The South Seneca Central School District has done some great work in the past concerning assessments, including intensive work on the Backward Design Process with Julie Kopp through LCI, Differentiated Instruction techniques, and Literacy Leader training through TST BOCES. We are seeking to strengthen our assessment system and are looking to develop strong and meaningful formative and summative assessments that will benefit our children and our everyday instruction. We are hoping that being more aware of our assessments will help us to become more informed about the information that our students are learning and are applying in their lives.

II. Guiding Questions.

- A. What types of assessments are our teachers using to determine the effectiveness of their instruction?
- B. Are teachers meeting the needs of all students during the feedback process?
- C. How do our given assessments affect the instructional decisions of the teacher?

III. Methodology

This audit was completed in collaboration with LCI (Learner-Centered Initiatives), facilitated by Jennifer Borgioli (of LCI), and in conjunction with TST (Tompkins-Seneca-Tioga) BOCES. There were many days of training for Liaisons at BOCES with Jennifer Borgioli. Then, Liaisons were expected to present the information that they learned about the process to their selected Grant Team at their home district.

A. Day one of LCI training at BOCES consisted of reviewing the foundations of quality assessment design and the components of such assessments. South Seneca was represented by the District Liaison.

B. Day two of LCI training at BOCES focused on what each of the entry points were and how we were to assess them in our own collected assessments. Again, South Seneca was represented by the District Liaison.

C. Preparation for Audit:

1. Liaison sent an email to the district requesting assessments that are used in the classroom. A total of 7 assessments were collected from two teachers. Our district has about 70 teachers total. Possible reasons for a low assessment return rate: request was submitted within two weeks prior to holiday break and/or hesitation to submit work (unknown to teachers how the information would be used).

2. Liaison met with members of the audit team. Liaison trained members of the audit team on rationale for the audit and what the entry points were. Audit team determined the focus of our work would be the Impact on Instruction entry point. This entry point contains three different areas: feedback, instructional decisions, and diverse

learners. We also agreed to send out another email to request assessments, addressing the hesitation that we were sensing from our colleagues.

3. Liaison and team members worked collectively on an email to send to the staff requesting assessments. Email was sent out and the response was much better. An additional 37 assessments were collected, bringing our total to 44 assessments collected across the district with approximately 35 teachers in the district participating.

4. A team meeting was conducted to review the entry point rubric and to score three sample assessments as a team. This meeting was designed to familiarize all members with the process and to assure inter-rater reliability while scoring the district assessments.

D. Assessment Audit

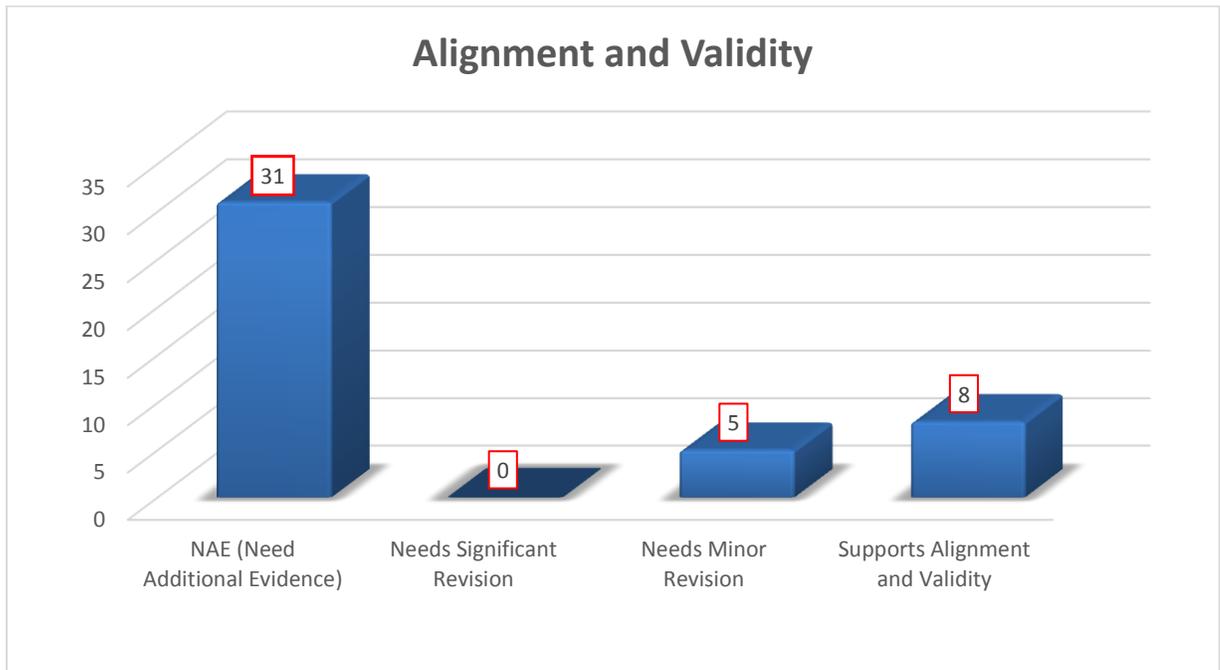
1. The team met over two sessions to carry out the assessment audit. Team members worked in pairs to review assigned entry points. The team decided to evaluate the first seven entry points, including all of the required entry points and two non-required entry points: Alignment/Validity, Impact on Instruction: feedback, instructional decisions, and diverse learners, Diversified and Balanced-types, differentiated, purpose, Authenticity, Thinking Demands and Rigor, and Reliability. The one entry point that we did not end up evaluating was Quantity due to the lack of information that we had. In order to effectively evaluate this entry point we would have to have a comprehensive report of all assessments given for each class and at each grade level.

2. After completing the assessment audit, results were compiled onto a graphic organizer that allowed us to see where our strengths and weaknesses were. Recommendations were then made based on our guiding questions and a conclusion was drawn.

IV. Data Analysis.

We were able to collect a total of 44 district assessments from the following disciplines: Science, Social Studies, ELA, Math, Spanish, Music, and Art.

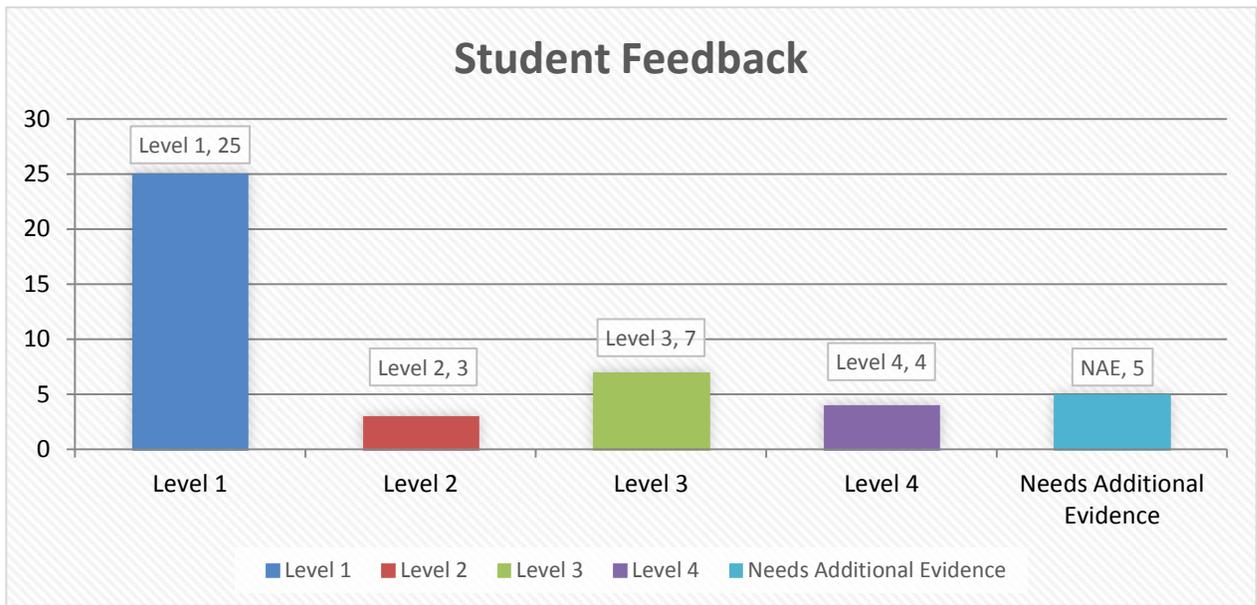
- A. **Alignment and Validity**: the degree to which the assessment is congruent with intended outcomes and standards. Our guides indicated that we must have the blue print or test map associated with the assessment in order to completely determine alignment to the standards.



The results of our investigation of this entry point show that many of our assessments did not include test maps or blue prints, which caused us to not be able to accurately assess our alignment and validity. Therefore, the majority of our assessments could not be evaluated for alignment to the standards. Out of 44 assessments, 31 of those assessments we deemed needed more evidence to be able to determine if the assessment was aligned to standards. This meant that there were no standards listed on the assessment and/or there were no test maps or blueprints included. 5 out of the 44 assessments needed minor revision. There was a connection to the standards revealed directly on the assessment, but the assessment did not line up directly with the standard or the standard given was not fully addressed. Finally, 8 out of 44 assessments supported a systematic attention to the alignment and validity entry point. To meet this level, assessments had to reveal a tight connection to the standards and had to show that the outcome of the assessment directly related to the standards.

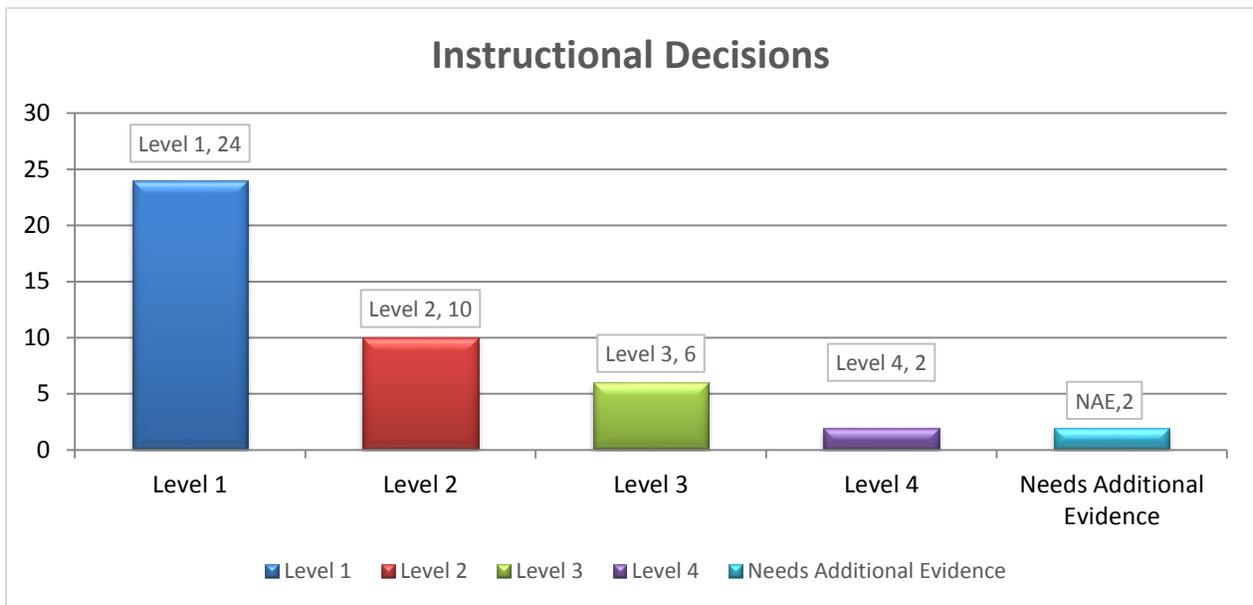
B. **Impact on Instruction** **This was our focus Entry Point.

1. **Student Feedback:** the way in which feedback is provided to the students during the assessment process. Evaluators looked at the type of feedback that would be given to the students (grade, written feedback, feedback identifying strengths and weaknesses, feedback that identifies strengths and weaknesses and questions for students to consider throughout the learning and assessment process).



The results for the Student Feedback entry point were spread over the four levels, but were very heavy at the lower end of the evaluation rubric. 25 out of 44 assessments were at Level 1, meaning that the assessment was mainly set up to give a score of some sort. There was no indication that any other type of feedback was to be given. 3 out of the 44 assessments were at Level 2, meaning there was evidence that a score and written feedback would be given. This was determined by a certain area marked for comments or feedback. 7 out of 44 assessments measured at Level 3. To be deemed a Level 3 the written feedback had to identify strengths and weaknesses. Finally, Level 4 contained 4 out of the 44 assessments. This level is very similar to Level 3, but the written feedback also provides questions for the students to consider throughout the learning and assessment process. Only 5 of 44 assessments needed additional evidence to determine what kind of feedback would be given.

2. **Instructional Decisions:** the way in which teachers use the results of the assessments to inform their instructional decisions.



Of 44 assessments, 24 were at Level 1 indicating that significant revision was needed in order to get on par with this entry point. To meet Level 1, the assessment determined who passed or failed, but the score is used just for that purpose. There may be feedback given, but the feedback is not helpful to the student. 10 out of 44 assessments measured at Level 2. These results are used to determine who needs extra help. The assessment may contain words such as “Check for Understanding.” 6 out of 44 assessments were at Level 3. Assessments in this level determine student strengths and needs and clusters students who share similar needs in to groups for future instruction. 2 out of 44 assessments scored at Level 4. These assessments determined strengths and needs and included information determining how results would be used to inform the development of learning experiences for the students that would address their needs and interests.

3. **Diverse Learners:** the degree to which assessment results are used to address the needs of diverse learners (including students with disabilities, English Language Learners, and gifted learners). How are teachers sharing the results with multiple stakeholders?

This Entry Point had only Level 1 and Level 2 results. It was difficult to assess this area because we felt that teachers may be practicing this more than their assessments indicated. Perhaps if there were lesson plans or memos included, we may have been able to do a better job determining a more accurate account of this Entry Point. 27 out of 44 assessments were at Level 1. This means that the assessment results are analyzed to determine who has failed an assessment. The information is shared with special education teachers so they can provide remediation or assistance to the student. 12 out of 44 assessments were at Level 2. This means that generic interventions may be used for these students that have been identified as struggling. We did not see evidence of Level 3 or 4 in the assessments and supporting materials provided. Level 3 and 4 at the Diverse Learners Entry Point would show learning activities developed to build on student strengths and weaknesses and individualized instruction plans that are designed to build on strengths and meet needs.

C. Diversified and Balanced

1. **Type:** assessments vary by type and may require student to recall information, create a product, demonstrate learning, or reveal thinking processes.

	Recall: Students are asked to recall or recognize the right answer.	Product: Students are asked to create a product to show what they know or can do.	Demonstrate: Students are asked to demonstrate or show what they know or can do.	Process: Students are asked to share the thinking processes they used while working on the assessment.
Number of assessments audited that contain the given type of assessment (some assessments contained more than one type):	34	27	3	10

It is important to note that many of the assessments contained more than one type of assessment. The majority of the assessments submitted had basic recall questions (either asking students to complete through multiple choice, matching, or fill in the blank). Recall was followed closely, however, by assessments that ask the students to create a product. In creating a product students are asked to show what they know or can do. This could include extended writing pieces, an art project, or solving complicated mathematical equations. Only 10 of the assessments asked students to share the thinking processes that they used while working on assessments. This included students completing math problems (showing their

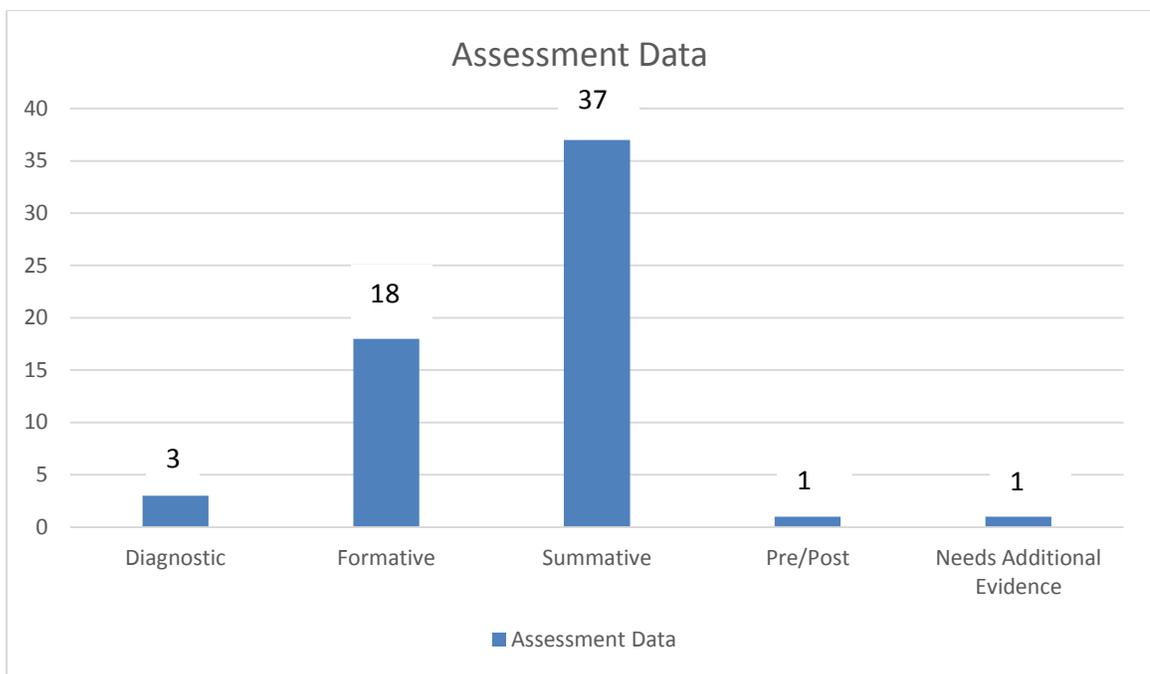
work and explaining why their answer is correct). Three assessments had students demonstrate what they know or can do.

Based on the assessments submitted, many of the assessments were determined to need significant or minor revision in order to support systematic attention to this entry point. 23 out of 44 assessments used an isolated type of assessment to determine evidence of student learning. This means that the assessment only contained one type of questioning (a multiple choice test or an essay, but not both). 21 out of 44 assessments contained multiple types of assessment. For example, multiple choice questions and short response questions. To achieve the highest level in this area, it was required that the assessments contain multiple types of assessment, but needed to include prompts that ask students about their own thinking. For example, “Which items on the assessment were easiest for you to answer? What made these items easy? Which items on the assessment were the hardest for you to answer? What made them difficult?”

2. ***Differentiated***: the degree to which assessments are modified to meet student needs and to allow students various ways to access content and /or demonstrate learning. Every single assessment that we scored was at the lowest level of this Entry Point assessment. Even though we determined that the assessments required students to perform in the same ways (with the exception of state mandated modifications for students with disabilities), I have a feeling that it is not an accurate representation of our assessments. However, in the assessments submitted, there was no evidence of planned differentiation.

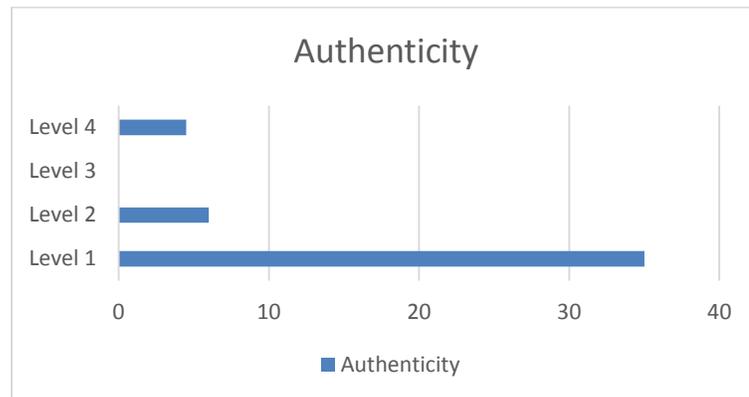
3. **Purpose:** refers to the timing of assessments.

Diagnostic	Formative	Summative	Pre/Post
Assessment occurs before instruction in order to inform instructional planning or placement.	Assessment occurs during instruction in order to inform immediate teaching and learning decisions.	Assessment occurs after instruction in order to summarize or evaluate student learning.	Assessment occurs before and after learning in order to measure or document growth.



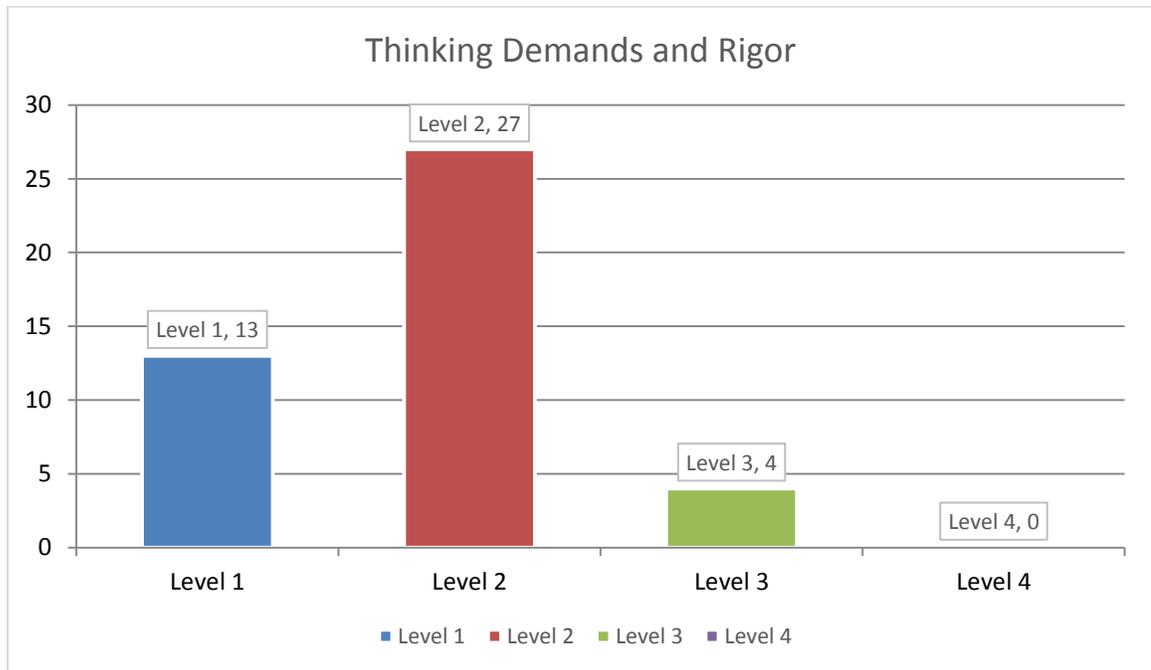
The majority of the assessments collected were summative and formative types. Some of the assessments overlapped as well (for example, a mid-unit test that would be considered summative, but is also formative because it will determine areas that the teacher may need to touch back upon in order to prepare the students for the end of the unit test).

D. **Authenticity:** the degree to which the assessment enables students to engage with tasks that address questions, issues or needs for an audience who could benefit from their work.



Many of the assessments submitted were lacking authenticity. 35 out of 44 assessments measured at a Level 1, meaning that the assessment was not connected to realistic problems or tasks. This would include a test or assessment just for teacher use (questions centered around a subject based area that have no relevance to the outside world). 6 out of 44 assessments involve students in plausible problems or tasks, but only includes the class or teacher as the audience. These tasks may be more relatable to the student's own life, but the teacher or their classmates may be the only ones viewing the results. 3 out of 44 assessments measured at Level 4. These assessments required students to engage in real problems and demonstrating or sharing their learning with audiences beyond school.

E. **Thinking Demands and Rigor**: the extent to which students can explore a concept, idea, issue, problem or question in depth. It supports deep thinking when students are able to investigate, analyze, synthesize, create, and apply their learning to new contexts or situations.



Many of the submitted assessments measured at a Level 1 or Level 2. This was also reflected in our test types. Level 1 shows that 13 out of 44 assessments were basic recall of isolated facts. Level 2 showed that 27 out of 44 assessments crossed over from just recall to including demonstrating understanding of factual and conceptual knowledge. This means that students were asked to use what they know about a subject and begin to apply it to a basic level (questions about scientific methods with open ended questions about the impact of a scientific discovery). 4 out of 44 assessments measured at a Level 3, meaning that in addition to the requirements of a Level 2, students were asked to analyze, create, and/or synthesize in response to teacher generated questions or problems.

F. **Reliability**: consistency. Based on our rubric, we needed to be able to see statistics of some sort.

This was an area that the submitted assessments were really lacking in. 42 out of 44 assessments measured at a Level 1, meaning that there was no evidence to support a claim of reliability. There were no statistics, or test result reports. 2 out of 44 assessments were at a Level 2. This means that reliability may be possible to obtain due to teachers collaborating to score, a Looking at Student Work Protocol, or evidence of a scanning system that could transfer information into statistical data. In order to score at a Level 4, assessments would have needed

to be accompanied with supporting statistics or rubrics would have needed to have anchors or exemplars.

G. **Quantity**: the number of assessments that students experience, take, or see over the course of a day, week, month, semester, or year.

This entry point was not required by the grant and was too difficult to accurately evaluate with the assessments that we collected. In order to do a quality job of assessing this entry point, we would need to gather much more information on assessments by grade level and subject area in order to complete a comprehensive review.

V. Recommendations

A. What types of assessments are our teachers using to determine the effectiveness of their instruction?

Based on the analysis of the assessments that were received it can be determined that assessments heavily rely on the basic recall of facts and are evaluated using an isolated type of assessment. There is evidence of students being asked to create a product showing what they know or can do, which is a step above basic recall. In order to move ahead in our assessment process it is recommended that teachers seek out ways to incorporate not only recall and product, but to include process and demonstration. Students need to be demonstrating what they know or can do and they must be able to share their thinking processes that they used while working on an assessment. Being able to SHOW their learning and REFLECT on their thinking processes is what will really help to firm up the knowledge that the students are learning in our classrooms.

B. Are teachers meeting the needs of all students during the feedback process?

To determine our effectiveness in this area, results from the Impact on Instruction—Student Feedback and Diversified and Balanced—Differentiated are being analyzed. Based on the results from the Student Feedback Entry Point, teachers are providing the students with a score as their main source of feedback. There are some exceptions with some assessments showing that feedback regarding strengths and weaknesses and questions for students to consider are evident, however, this was not apparent on an overwhelming majority of the assessments. It is important to note the reason for our assessments. Are we assessing to help the student to understand where they stand and how they can improve, or are we assessing just to get that number to put in the gradebook? It is recommended that we move more toward providing the student with written feedback that identifies their strengths and weaknesses, including guiding

questions for students to consider throughout the process to improve their educational experience.

Based on the Differentiated results, we are not reflecting any kind of differentiation in our assessments. This is an area that we need to refocus on. We have had work on differentiation in the past, but it is not currently reflected in the assessments that were submitted.

C. How do our given assessments affect the instructional decisions of the teacher?

The way in which teachers use the results of the assessments to inform their instructional decisions is not blaringly apparent in the assessments that were evaluated. This does not mean that it is not being done, but if it is being done, it needs to be better documented. This area ties in nicely with student feedback. Based on the feedback that we give students, we as teachers can be determining how to group our students to provide extra support. Many of our assessments reflected that assessments were only being used to determine who passed or failed with no indication of determining who needs extra help or how we will be providing that extra help. Assessment maps need to show our intent in using the results of the assessment data to move forward.

VI. Future Plans

A. Based on our audit, we have determined that out of the 44 assessments that we audited, 44 of the assessments will be kept, none will be modified or eliminated at this time, and we will create one new assessment. Our teachers are doing really great work in their classrooms. We do have areas that need improving, but through future half-days, conference days, and professional development offerings, we should be able to address many of these areas.

B. This audit and the action plan that came about as a result of our audit will be shared with the public on the www.southseneca.org website. Links to the audit will be shared in the school newsletters so that parents are informed of the report's presence on the website.

C. Assessment Design: A small team of teachers will work on creating an ELA 9 assessment. Design work time will occur primarily on March 10, March 11, April 17, and June 1 in conjunction with TST BOCES and LCI.

F. The South Seneca district will have opportunities for professional development in the spring of 2015 (Assessment Creation through LCI) and May and June of 2015 through TST BOCES (Expeditionary Learning "Student Engaged Assessment" workshop (3 days) and "Google

for Assessment” workshop (1 day). Beyond these given opportunities, our district will continue to work on assessment related areas in our professional development work.

VII. Conclusions

Although in the past years our district has done extensive work with Student Work Protocols, Differentiated Instruction, and the Backward Design Process, it seems that it is not being effectively reflected in our assessments. There are hints of the work there, but the work needs to be more apparent. The work has not been lost, but it is not being reflected in our assessments and our explicit plans for carrying out the assessments.

APPENDIX

Summary of Assessment Audit Results

Alignment and Validity

	✓-	✓	✓+	NAE
# of Assessments	0	5	8	31

Impact on Instruction-Student Feedback

	1	2	3	4	NAE
# of Assessments	25	3	7	4	5

Impact on Instruction-Instructional Decisions

	1	2	3	4	NAE
# of Assessments	24	10	6	2	2

Impact on Instruction-Diverse Learners

	1	2	3	4	NAE
# of Assessments	27	12	0	0	5

Diversified and Balanced-Type

	✓-	✓	✓+
# of Assessments	23	21	0

	Recall	Product	Process	Demonstrate
# of Assessments	34	27	10	3

Diversified and Balanced-Differentiated

	✓-	✓	✓+
# of Assessments	44	0	0

Diversified and Balanced-Purpose

	Diagnostic	Formative	Summative	Pre/Post	NAE
# of Assessments	3	18	37	1	1

Authenticity

	1	2	3	4
# of Assessments	35	6	0	3

Thinking Demands and Rigor

	1	2	3	4
# of Assessments	13	27	4	0

Reliability

	1	2	3	4
# of Assessments	42	2	0	0