# Washington Mathematics Assessment Handbook

**Introduction**

The purpose of this project is to design a comprehensive assessment system for mathematics standards that will: 1) focus instruction across a school and/or a district on specific areas of knowledge and skill referred to as “measurement topics” and 2) provide a vehicle for teachers and/or other school staff to keep track of the progress of individual students on each measurement topic using teacher-designed formal and informal assessments. To these ends, the Washington State ESD Regional Mathematics Coordinators identified a set of mathematics measurement topics for each grade level based on the Washington state mathematics standards. These topics are listed in the chart below.

# K-5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Measurement Topic** | **K** | **1** | **2** | **3** | **4** | **5** |
| **Strand: Geometry/Measurement** |  |  |  |  |  |  |
| Perimeter, Area, Surface Area and Volume |  |  |  | + | + | + |
| Properties of Geometric Figures | + | + | + | + |  | + |
| Lines/Angles |  |  |  | + |  | + |
| Money |  |  | + |  |  |  |
| Measurement Using Non-Standard units/location terms | + | + | + |  |  |  |
| Time |  | + | + |  | + |  |
| Measurement using Standard and Metric Systems – length, weight, mass, capacity |  |  | + | + | + |  |
| Temperature |  |  |  | + |  |  |
| Proportionality and similarity |  |  |  |  | + |  |
| Transformations |  |  |  |  |  | + |
| **Strand: Number** |  |  |  |  |  |  |
| Place Value |  | + | + | + | + |  |
| Counting/Naming | + | + | + | + |  |  |
| Comparing/Ordering | + | + | + | + | + |  |
| Rounding |  |  |  | + | + |  |
| Fractions |  |  | + | + | + | + |
| Multiple Representations of fractions, decimals and the number system |  |  |  | + | + | + |
| Factors |  |  |  |  | + | + |
| Number representations (Exponents, Scientific Notation) |  |  | + |  |  |  |
| Number Systems |  | + |  |  |  | + |
| **Strand: Operation** |  |  |  |  |  |  |
| Addition and Subtraction |  |  | + | + |  |  |
| Compose and Decompose Numbers | + | + |  |  |  |  |
| Alternate and Mental strategies |  | + | + | + | + | + |
| Estimation |  |  | + | + | + | + |
| Multiple representations/Models | + | + | + | + | + | + |
| Performing operations using fractions and other components of the number system |  |  |  |  |  | + |
| Converting between fractions, decimals and mixed numbers |  |  |  |  | + |  |
| Multiplication and Division |  |  |  | + | + | + |
| **Strand: Algebra** |  |  |  |  |  |  |
| Patterns | + | + | + |  |  | + |
| Writing/Evaluating Expressions, equations and inequalities |  | + | + | + | + | + |
| Graphing |  |  |  |  | + | + |
| **Strand: Data Analysis, Statistics and Probability** |  |  |  |  |  |  |
| Data Displays |  |  | + | + | + | + |
| Analysis of data |  | + |  |  |  |  |
| Measures of Variability |  |  |  |  | + | + |
| Theoretical and Experimental Probability |  |  |  |  | + |  |
| **Strand: Processes** |  |  |  |  |  |  |
| Problem Solving | + | + | + | + | + | + |
| Communicating the results of a problem | + | + | + | + | + | + |
| Mathematical experimentation |  |  |  | + | + | + |

At each grade span, the measurement topics are presented in a rubric format as shown in Figure 1.

Figure 1 depicts the 8th grade rubric for the measurement topic Theoretical and Experimental Probability. To understand this format, consider score 3.0. It states that while engaged in tasks involving the Theoretical and Experimental Probability, the student will:

8.3.F determine probabilities for mutually exclusive, dependent, and independent events for small sample spaces

This score 3.0 statement represents the expected behavior for 8th grade students regarding the measurement topic of Theoretical and Experimental Probability. That is, score 3.0 in each rubric represents the “target level” of knowledge for a measurement topic at a given grade level.

The sample task portion of the rubric (the right side) presents possible assessment tasks that a teacher could use to assess a student’s current level of knowledge regarding this topic. In this case, the rubric suggests that the teacher could ask the student to solve an example of a probability word problem by determining if a given probability is correct. It is important to keep in mind that these tasks are only examples. The teacher is free to develop their own task to measure student progress on the expectation.

***Figure 1: Rubric for the Measurement Topic of Theoretical and Experimental Probability at grade 8***

|  |
| --- |
| **Strand: Data Analysis, Statistics and Probability** |
| **Topic: Theoretical and Experimental Probability** |
| **Grade: 8** |
| **Score 4.0** | **In addition to Score 3.0, in-depth inferences and applications that go beyond what was taught.** | **Sample Tasks** |
|  |
|  | **3.5** | In addition to score 3.0 performance, in-depth inferences and applications with partial success. |  |
| **Score 3.0** | **The students are expected:**

|  |
| --- |
| * 8.3.F determine probabilities for mutually exclusive, dependent, and independent events for small sample spaces
 |
|  |

**The student exhibits no major errors or omissions.** |  |
|  | **2.5** | No major errors or omissions regarding 2.0 content and partial knowledge of the 3.0 content |  |
| **Score 2.0** | **There are no major errors or omissions regarding the simpler details and processes as the student:** * recognizes or recalls specific terminology such as:
	+ probability, events (mutually exclusive, dependent, and independent)
* performs basic processes, such as:
	+ matches given probabilities to various events (mutually exclusive, dependent and independent)

**However, the student exhibits major errors or omissions regarding the more complex ideas and processes.** | * Ask the student to identify or produce definitions to given terms (e.g., probability, events (mutually exclusive, dependent, and independent))
* Ask the student to match given probabilities to descriptions of various events
 |
|   | **1.5** | Partial knowledge of the 2.0 content but major errors or omissions regarding the 3.0 content |  |
| **Score 1.0** | **With help, a partial understanding of some of the simpler details and processes and some of the more complex ideas and processes.** |
|  | **0.5** | With help, a partial understanding of the 2.0 content but not the 3.0 content |
| **Score 0.0** | **Even with help, no understanding or skill demonstrated.** |

Next consider score 2.0. In effect, it states that the student cannot determine probabilities, but can recognize or recall correct examples of probabilities. The sample task in this case suggests that the student could match probabilities to descriptions of events. In addition, Score 2.0 asks the student to recognize or recall the definitions to important vocabulary necessary to understanding score 3.0. In this case it includes the terms probability and events. The sample task asks the student to identify (recognize) or produce (recall) definitions to these terms.

Score 2.0 behaviors are obviously more basic and simpler than the score 3.0 behaviors. That is why they are referred to as the simpler “details and processes” in the rubric. Score 2.5 is between scores 3.0 and 2.0. It indicates that the student can perform the score 2.0 behaviors successfully but is only partially successful at the score 3.0 behaviors. For example, a student can match probabilities to events and recall the definitions, but is only able to determine the probability in one of the situations described in score 3.0. Score 1.5 indicates that the student is partially successful at the score 2.0 behaviors.

Score 1.0 indicates that without help the student is unable to successfully demonstrate a particular skill. However, with help from the teacher (e.g., direct guidance, suggestions, prompts) the student has partial success with the score 3.0 and the score 2.0 behaviors. Score 0.5 indicates that with help the student has some success with the score 2.0 behaviors but still has no success with the score 3.0 behaviors. Score 0.0 indicates that the student has no success even with the teacher’s help with either the more complex (i.e. score 3.0) or the simpler behaviors (i.e. score 2.0) for the measurement topic.

Finally, score 4.0 indicates that the student not only demonstrates all behaviors indicative of scores 2.0 and 3.0, but also makes inferences and applications that go beyond what was taught in class.. A student could meet the 4.0 expectation by demonstrating any learning that goes above and beyond what was explicitly taught, and demonstrating that knowledge in a variety of ways. For example, a student might be asked to apply their knowledge of probability in a new and unique way or in a new and unique situation that was not taught. A student who could apply probability in a way that was not taught, in addition to successfully completing the score 3.0 and the score 2.0 tasks, would be assigned score 4.0. The inclusion of Score 4.0 in the rubric is simply intended to illustrate what a student *might* do, not what they **must** do.

Score 3.5 also indicates that the student demonstrates all score 2.0 and 3.0 behaviors. Additionally, the student attempts to do things like explain the reasons that objects hold their position in space, however, the student is only partially successful in these attempts.

**Teachers Designing and Administering Their Own Assessments**

The system of rubrics provided in this document is intended to be used by teachers to design, administer, and score their own classroom assessments. Assessment in this context is intended to go beyond a typical test. Marzano Research Laboratory (MRL) defines assessment as “anything that a teacher does to gather information about a student’s knowledge or skill regarding a specific topic.” Anything that the teacher does in the classroom including learning activities, homework, class work, tests, quizzes, projects, presentations, group work, etc. can be used to collect data about a students’ progress towards meeting the standard. MRL breaks assessment into three general types: Obtrusive, Unobtrusive and Student Generated.

**Obtrusive Assessment:**

Obtrusive assessments include a wide variety of classroom activities and assessments, all designed and used to assess student progress towards demonstrating success on the measurement topic. The key to an obtrusive assessment is that the teacher intends for the given activity or test to be an assessment, and generally, the student knows that they are being assessed. Instruction stops in order to assess student progress. Generally, anything that a teacher does in the classroom can be used as an obtrusive assessment, given that the intention is to use it as an assessment. Remember that an obtrusive assessment becomes an assessment portion of the school day, not instruction time.

Typically, the teacher will take a grade or collect a score when the student is given an obtrusive assessment. This data is collected and used to make judgments about student progress (formative assessment) or to help determine a final grade (summative assessment).

**Unobtrusive Assessment:**

Unobtrusive assessments are the opportunities that a teacher has to observe a student performing a skill or demonstrating their knowledge of a measurement topic. These assessments occur when the teacher becomes a passive observer of the learning environment, watching the student to determine level of skill. The learning activity does not stop so that the student can be assessed, rather the teacher watches the activity and collects and records data without interrupting the activity or learning. Essentially, the teacher watches a student as he/she performs a skill or demonstrates knowledge in order to discern ability level.

Unobtrusive assessments work very well for observing student demonstration of procedural knowledge. For example, a teacher can readily observe a student or class performing the steps of a procedure and record data about student understanding of and ability to perform the procedure.

Unobtrusive assessments can be used to observe declarative knowledge as well. By noting how a student interacts with their peers around the desired knowledge, the teacher can gather data about a student’s current level of understanding.

Unobtrusive assessment data can be collected just like obtrusive data. This data is just as valid and reliable and can be used for formative or summative assessments purposes.

**Student Generated Assessment:**

Perhaps the best use of the rubric is to turn the responsibility for assessment over to the student. With a well –written rubric, a student can determine how they are going to prove their skill or knowledge to the teacher. In this type of assessment, the teacher encourages the student to develop their own assessments based on the rubric. Once the assessment is completed and scored, the teacher can use the data from this assessment just like any other assessment data.

In order for a teacher or student to use a student-generated assessment, the student should initiate the request to “prove” to the teacher that they are capable of a better score then the teacher has currently given them. The student will decide what form of assessment they would like to do in order to demonstrate their current level. The teacher should provide guidance to make sure that the assessment will meet the requirements of the scale. Keep in mind that student generated assessments could be used for data at any point in the formative assessment process. In addition, the student should be encouraged to use any form of assessment. Ideally, the student will choose assessment options that fit their particular strengths as a student. This increases student motivation and engagement.

Once the student has chosen the assessment option, they are on their own to complete it. When it is completed, the teacher will score the assessment using the same criteria (scale) as any other assessment, and record the data. This data can then be used to determine the overall trend of student learning like any other teacher generated assessment.

**Assessment Approaches**

There are two basic approaches to designing assessments using these rubrics: 1) designing assessments that include all levels of the rubric and 2) designing assessments that include one level of the rubric only. It is also possible to design an assessment that allows the teacher to collect data on multiple rubrics at the same time.

**Assessing All Levels of the Rubric**

To illustrate this approach, assume that an 8th grade teacher wishes to design an activity to assess the measurement topic Theoretical and Experimental Probability. The teacher would make sure they have a part of an activity or items that address score 2.0. From the rubric in Figure 1 we see that score 2.0 involves matching probabilities to a description of an event. To assess score 2.0 the teacher might provide students with a set of descriptions and a list of probabilities and ask them to match the probability with the appropriate description. Successfully completing this task would indicate that a particular student had obtained score 2.0 status.

To assess score 3.0, the teacher might ask the student to solve problems that involved determining the probability. Successful completion of this task would indicate that the student has reached score 3.0 status.

Finally, to assess score 4.0 status the teacher would include an item that goes beyond determining the probability for small sample spaces. For example the teacher could ask students to go one step further by determining the probability for large sample spaces, in a new and unique situation, or otherwise apply their knowledge of probability in a way that goes above and beyond what was explicitly taught. A student who could apply probability in this way, in addition to successfully completing the score 3.0 and the score 2.0 tasks, would be assigned score 4.0. Remember that the example just cited is one of many possible ways a student might achieve score 4.0. The teacher and student have the freedom to demonstrate score 4.0 in any way that they see fit.

The example above involves a classroom assessment assigned to the entire class. The 8th grade rubric for Theoretical and Experimental Probability can just as easily be used with individual students. For example, the teacher might call a particular student up to her desk while other students are working independently. The teacher would ask the student to match given probabilities to given situations (Score 2.0). Next, the teacher would ask the student to determine the probability (Score 3.0). Lastly, the teacher could ask the student to use probability in a way not explicitly taught (Score 4.0). The student’s performance on each of these three interactions would determine the student’s score for this assessment. In short, the rubrics can be used by classroom teachers to design, administer, and score a variety of types of classroom assessments for the measurement topics.

**Assessing One Level of the Rubric Only**

Some teachers prefer to design and administer assessments that address one level of the rubric only. For example at the beginning of a unit, assessments include items, tasks or activities for score 2.0 only. In this case the first assessments in a unit would focus on recognizing or recalling the position of objects in the universe. Once students had demonstrated competence in this score 2.0 content, the teacher would focus on score 3.0 content and so on.

In this system, students are assessed only on what has been or is being taught in class at a particular moment in time. Typically students are more familiar and more comfortable with this approach. However, the drawback of this approach is that it doesn’t provide students with a picture of where the unit is going and what will be expected of them by the end of the unit.

**Assessing multiple topics using one assessment**

Another option is to design an assessment with the intention of measuring multiple math topics. For instance, a unit of study might include several related topics that are taught concurrently. Teachers can design assessments that measure all of these topics at the same time. It is important to keep the scores of each topic separate from each other to ensure accurate measurement. Teachers should have a great deal of flexibility to design assessments that address single or multiple topics.

**Students Keeping Track of Their Own Progress**

One of the simplest and most powerful uses of the measurement topics is to provide students with venues for keeping track of their own progress. To demonstrate, consider Figure 2, depicting a specific student’s scores on the measurement topic of Theoretical and Experimental Probability.

***Figure 2: Student Progress Chart***

|  |
| --- |
| **Keeping Track of my Learning**Name: I. H. Measurement Topic: Theoretical and Experimental Probability My score at the beginning: 2 My goal is to be at 3 by Nov. 30th Specific things I am going to do to improve: Work 15 min. three times a weekDate assessment data collecteda Oct. 5th f Nov. 26th b Oct. 12th g c Oct. 20th h d Oct. 30th i e Nov. 12th j  |

In this example, the student is tracking his/her progress by recording six scores which were gathered from formal or informal assessments constructed by the teacher. The first assessment was on October 5, the next was on October 12 and so on. Each student in the class would begin with a blank form similar to the one filled in as an example above. After each assessment scored by the teacher each student would record his or her score on their individual progress chart. This not only provides students with a visual representation of their progress but invites discussions between teacher and students about areas of strength and weakness along with strategies for improvement as well. It also provides students with opportunities to set academic goals.

The data from individual progress charts can also be aggregated to obtain a snapshot of the progress for an entire class or set of classes. To illustrate, consider the example shown in Figure 3.

***Figure 3: Aggregated Student Data***

|  |
| --- |
| **Recording Student Achievement – Multiple Classrooms**Teacher Name: Ms. Pickering Measurement Topic: Theoretical and Experimental ProbabilityClass/Subject: Math Grading Period/Time Span: Quarter 2Date assessment data collected1 Nov 2nd 6 2 Nov. 15th 7 3 Dec. 5th 8 4 Dec. 15th 9 5 Jan. 6th 10  |

This figure depicts scores from a number of classes that involve 110 total students. It shows how many students are above a specific “performance standard.” For example, if the school or district has established that it wishes all students to achieve a rubric score of at least 2.0 on the rubric for a particular measurement topic, then the score of 2.0 is the performance standard for that particular topic. For the measurement topic of Theoretical and Experimental Probability 50 percent of the students were at or above the performance standard on November 2 as they were for the next two assessments. However, by December 15, 70 percent of students were at the performance standard or above.

This type of aggregated data can provide teachers and administrators with a snapshot of progress for a specific math measurement topic at a number of levels: classroom, grade level, school, or district. Such aggregated data can be the basis for identifying future instructional emphases for teams of teachers. If the aggregated data indicates that an insufficient percentage of students in a particular grade level are at or above the designated performance standard, it would suggest that teachers, student support staff, and administrators work together to determine how best to enhance student progress for that particular measurement topic.

**As the Basis for Grading and Reporting Final Status**

Measurement topics can be used as the basis for classroom grading in traditional or nontraditional ways. To illustrate, consider an 8th grade teacher who is attempting to keep track of students’ performance on four pre-determined measurement topics in Math within a given quarter:

* Theoretical and Experimental Probability
* Data Collection
* Measures of Variability
* Analysis of Data

She would utilize four rubrics for those measurement topics each modeled after the rubric template in Figure 1.

Over a nine-week period of time (let’s say), the teacher would administer a number of assessments on these measurement topics. Some of these assessments might take the form of quizzes and tests. Some forms of assessment also could be quite informal in nature, such as an interaction the teacher has with a specific student. A teacher could use different types of assessment, such as obtrusive, unobtrusive or student generated. Figure 4 depicts the scores for four measurement topics that have been addressed over a given grading period. Two of these topics might have been addressed during one unit of instruction, and the other two during a different unit.

***Figure 4: Rubric Scores for a Student on Four Math Measurement Topics***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Assessment #** | Theoretical and Experimental Probability | Data Collection | Measures of Variability | Analysis of Data |
| 1 | 1.5 | 2.0 | 1.0 | 1.0 |
| 2 | 2.0 | 2.5 | 2.0 | 2.5 |
| 3 | 3.0 | 3.0 | 3.0 | 3.0 |
| 4 | 3.0 | 3.0 | 4.0 | 3.5 |
| **Final Score** | **3.0** | **3.0** | **4.0** | **3.5** |

Figure 4 contains the scores for a single student on four measurement topics. Four assessment scores have been recorded for each topic. These assessments might have been specific to a single topic or they might have included two topics. For example, the teacher could have given a single assessment that included the topic of Data Collection and the topic of Analysis of Data. Rather than assign a single rubric score on this assessment, two rubric scores would be assigned—one for each topic. On another occasion the teacher might have designed an assessment that addressed a single topic like Measures of Variability. One score would have been assigned to this assessment because it addressed one topic only.

The row in Figure 4 entitled “final score” reports the final status for the student in each topic. **It is important to note that this final score is not obtained by simply averaging the previous scores.** Rather the teacher would look at the trend of scores over time and assign a topic score that most accurately represents the student’s status at the end of the grading period. To generate this final score the teacher would use a variety of information including:

* the formal assessments that have been administered and recorded (i.e. the scores shown in Figure 4)
* information obtained from informal data gleaned as a result of interacting with the student in class and observing the student
* the trend data from a computer grading program

We recommend that teachers draw from a variety of data sources to make a well informed decision about a student’s final status for a given grading period on each measurement topic. The average is rarely appropriate for a final score on a given topic. To illustrate, consider the four scores in Figure 4 for the topic of Measures of Variability: 1.0, 2.0, 3.0, and 4.0. The average of these scores is 2.5. But this score does not reflect the student’s true level of learning at the end of the grading period. Even though the student began with a score of 1.0, she had demonstrated level 4.0 competence by the end of the quarter.

To construct an overall grade for Math for a given 9 week period of time (let’s say), the final scores on each measurement topic would be averaged in some weighted or un-weighted fashion. Averaging final topic scores might sound like a contradiction to the statement above that the final score for a given topic should not be computed by averaging scores on previous assessments for the topic. This later statement is true; you shouldn’t average scores over time for assessments within the same measurement topic. This is because students should be “getting better” regarding the understanding of or skill at a given measurement topic over time. An average wouldn’t accurately reflect their level of understanding or skill at the end of the grading period. In fact averaging will underestimate their true status because it doesn’t take into account the fact that a student has improved over time. A student could start a grading period with a score of 0.0 on a given topic and end the grading period with a score of 4.0. However when you have to put together final scores **across** measurement topics, you have little option but to do something like average them. This is because the different topics measure different types of information and skill. By definition, whenever you aggregate final scores on different topics you are creating an overall score across different things. The best you can do in this case is to use a weighted or unweighted average. The only other option is to devise a system that requires students to receive specific minimum scores on specific topics. For example, to receive a grade of A the student might have to obtain a minimum score of 3.0 on two specific topics and a minimum score of 2.0 on two specific topics that have not been fully addressed during that grading period (but will be fully addressed during the next). In short, there are many ways to construct a final grade but it is always important to remember that an overall grade that cuts across multiple measurement topics will force the teacher to use some artificial and arbitrary way to combine scores on different types of information and skill.

If a weighted or un-weighted average is used, it could be readily transformed into traditional A, B, C, D and F grades. Figure 5 provides a sample of a possible grading scheme.

***Figure 5: Conversion Scale to Traditional Grade***

|  |  |
| --- | --- |
| **Average Rubric Score across Multiple Topics** | **Traditional Grade** |
| 3.00-4.00 |  | A |
| 2.50–2.99 |  | B |
| 2.00-2.49 |  | C |
| 1.50-1.99 |  | D |
| Below 1.50 |  | F |

Using final topic scores, any number of measurement topics addressed within a grading period can be combined to form an overall grade based on the traditional A,B,C,D,F scale. Of course, this is an example only. The school district or individual schools must design a system that best reflects their formative and summative assessment needs.

At the elementary level it is common to use a reporting system other than the traditional A, B, C, D, F format. For example, the format in Figure 6 might be used instead:

***Figure 6: Conversion Scale to Nontraditional Grading Scale***

|  |  |
| --- | --- |
| **Average Rubric Score across Multiple Topics** | **Traditional Grade** |
| 3.01-4.00 |  | Meets with Excellence |
| 2.50–3.00 |  | Meets with Proficiency |
| 1.50-2.49 |  | Approaches Proficiency |
| Below 1.50 |  | Well Below Proficiency |

Again, this is an example only. The school district or individual schools must design a system that works best for them.

**Commonly Asked Questions**

When first using a system of measurement topics and scoring scales, like the one presented in this document, teachers will commonly have many questions. Four common questions are addressed here.

**How Many Assessments Do I Need to Assign a Final Score on a Measurement Topic?**

In Figure 4, four scores are reported on each measurement topic. It is useful **but not necessary** to have four scores per student per topic. The general rule to follow is that a teacher should have **as many** or **as few** scores as are necessary to arrive at a final score that the teacher believes accurately reflects a student’s status at the end of a grading period. For some students a teacher might be sure of a particular student’s status after a few assessments only. For some students a teacher might need more than four assessments.

**What Counts as an Assessment?**

In a system like the one presented in this document, an assessment can take many forms. Certainly traditional paper/pencil tests are a common form of assessment. However, an informal discussion with a particular student on a particular topic can also be a valid way to collect assessment data. Given this flexibility it is fairly easy for a teacher to collect multiple scores for a topic. If the teacher believes that she needs more information to determine a student’s final status she can ask that student to meet individually with her and use the subsequent discussion as a type of assessment, or the teacher can ask that students to complete some extra exercises.

**How Can I Make the Rubrics “Student Friendly?”**

The rubrics in this document are written for teachers and other school staff. They can and should be translated into student language. This is done by presenting the class with a rubric and as a group rewording the score 2.0, 3.0, and 4.0 content in a way that gives students specific examples of what constitutes a score 2.0 response, a score 3.0 response, and a score 4.0 response.

**Where Does Student Self-Assessment Fit In?**

If rubrics have been rewritten in student friendly language, they can be used by students to score themselves on assessments. The scores students assign themselves can then be compared with the scores the teacher has assigned on a specific topic. This should facilitate some powerful dialogue between teacher and student regarding the scores that most accurately represent his or her status at any point in time.