

Everett Public Schools
4-Trait
Writing Assessment
Scoring
Guidelines:



*Many people hear voices when no one is there. Some of them are called mad and are shut up in rooms where they stare at the walls all day. Others are called writers and they do pretty much the same thing.
~Meg Chittenden*

"Writer in Control" Framework

Writer is in Control

(Range from 3.5 to 4)

The writer controls what is on the page. Strengths clearly outweigh the weaknesses, if the latter are present at all. Writing skills are evident, not over done, and the reader does not have to infer the writer's intent or the direction the writer is going. Writer organizes. Writing communicates clearly. Writer is competent. Writing is more than functional. It is significant in its, for example, style or content. Writer does the work and leads the reader.

Writer is Wobbling Between Control and Non-Control

(Range from 2.5 to 3.5)

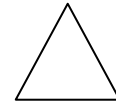
The writer is not fully in control, yet the writing communicates. Writing strengths are vying with needs for writing improvement. The writing needs revision, editing, but at the same time demonstrates more than an emerging level of writing skills. Writing serves as a functional level, communication tool, but the reader wants more out of style, content, and the use of appropriate organization, sentence, and convention strategies. Writer, at times, leads and, at other times, hands over to the reader the lead for some level of understanding or implications.

Writer is Yet to Be in Control

(Range from 1 to 2.5)

The writing process overpowers the writer. Need for improvement overshadows the writing strengths. Skills are not sufficiently present or are emerging with hints of the writer's intent. Usually the reader needs to infer or decode to understand the meaning. Writing is barely, if at all, functional for communicating ideas, evidence and affect. Reader does much work to understand the writer.

D.V.S. Error Ratios



Density

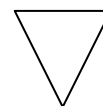
Weigh the proportion of errors to the amount of writing done well. This includes the ratio of errors to length.

Variety

Weigh the proportion of errors that range across indicators against the proportion of errors that remain within an indicator.

Severity

Weigh the ratio of "basic" errors in relation to "higher level" (grade level) errors.



Scoring Strategies

Not in Control

Conventions
(start with 2– work up or down)



Organization



Style



Content

Scoring Strategies

Wobbling or In Control

Organization
(start with 3– work up or down)



Style



Conventions



Content

Scoring Session

Tips and Strategies

Personal Scoring Pitfalls

Personal preferences can affect a reader's perception of a writer's response. By being aware of and monitoring our work, we can avoid impacting the scores by these preferences.

Correct Response Myth

For many classroom tests, writing is scored correct or not, like in the case of the essay prompt "What were the causes of the Civil War?" In writing assessment, however, this is not the rule. Writers are **free to respond in any way they choose as long as it is an effort that is on-topic**, and in some cases, in the correct mode. In short, there are no "right" answers. The response need only be a reply to the prompt.

Appearance Impression

Handwriting - Readers are often swayed by how a paper looks on the page—by its neatness and the appearance of the handwriting. The fact is, though, superior papers (with respect to ideas, style, and organization) can look sloppy and can be hard to read; on the other hand, relatively mediocre papers may be very neatly presented.

Short-Long Papers: The length of a writer's paper is not part of the scoring criteria, unless it has too few sentences to rate, usually three. Size of a writer's handwriting can make a paper look longer or shorter on the page than it actually is.

Short, simple papers beyond grade four probably don't show proficiency performance in most traits. And then again, longer papers can be over penalized because more mistakes are apparent. Longer papers should be viewed as having proportionately the same number of mistakes and, therefore, should potentially receive the same scores as a short paper. The goal of good writing is conciseness, not length. Essays should say it and be done.

Emotions

Empathy-Creativity-Unorthodoxy Impression: Like a paper's appearance, if the writer evokes strong emotions in the reader, either positive or negative, the paper's style/voice trait(s) will often drive the rating of other traits. Papers that engage the reader with positive emotions will result in the overlooking of errors in other traits. Conversely, very creative or unorthodox content, organization, or sentence structure may often confuse or turn off the reader and result in lowered scores in other traits. In general, read Style traits early and then ignore the emotions or the unorthodoxy when rating the other traits.

Positive-Negative Leniency Error: Some readers tend to be too hard or too easy on everybody—almost as a kind of personal belief system or policy. All points/levels of the writing scoring guide should be useful for and used by every reader. You should have high and low scores across students and your ratings of their performance should be distributed approximately like other raters.

Offensive or Disturbing Content: Writers may be vulgar, sexist, racist, violent, or adopt an "anti-whatever" point-of-view. They may also be naive or take a narrow point of-view compared to the reader's experiences. Or they may come across as brash, sassy, cute, self-aware, shy, surly, or honest.

The reader can not let his or her lifestyle or maturity level influence the ratings of the writer's ideas, either positively or negatively. Know yourself. If you find the content troubling and think you can not rate it without bias then pass it to another reader.

The Sympathy Score: Readers are sometimes reluctant to give low ratings because they view these scores as penalties, because the writing's content causes the reader to sympathize with the writer, or the reader believes the prompt was "really bad." Assessment cannot do its job to identify the extent to which a student's writing performance is strong or weak in a given area if the results are truncated and do not span the full range of the scoring guide's scale.

True enough, the student who receives low scores may find that depressing. On the other hand, if everyone receives relatively average or high scores, the scores and subsequent feedback becomes meaningless. The efforts of the student will have been wasted. Students and parents who sense that scoring is not an accurate reflection of the student's writing will not respect the scoring system or the advice based upon it. Research indicates that students prefer writing assessments that match their own expectations, self-assessments, and judgments about their performance—regardless of whether such scores turn out to be high or low. Moreover, few if any of these scores are used by teachers as a part of their students' grades.

General Scoring Pitfalls

Common Language Based on Common Examples: Readers often assume that everyone else has the same definition in mind when they are judging writing performance. For example, Organization in one scoring system may include attributes of Sentence Fluency in another system. Only if the readers have worked together to clarify definitions, examples, and meanings of the writing trait(s) does the writing assessment scores have any meaning beyond the reader him/herself. In short, for students, parents, teachers, or other colleagues, know the scoring guides thoroughly and differentially.

Double Jeopardy: A few readers tend to be too hard or too easy on a given trait—often, it's Conventions. For example, some readers are so bothered by poor spelling that they find it hard to score the paper relatively high if it contains spelling errors, regardless of its other merits. Readers should not penalize the writing as a whole if the same error occurs more than once. Simple editing will correct simple mistakes. If a writer has major problems in one trait, rate that trait first and, then, ignore those errors in rating the other traits.

Forced-Choice Dilemma: Readers may consistently rate either higher or lower than they feel the quality of the paper merits simply because they have no mid level on which to place the paper. A reader faced with this predicament will rate the paper based on his or her personal philosophy, tending to be a little easier on papers than seems warranted, or a little harder, whichever seems less offensive. Again, the solution is the use of the anchor papers in relation to the verbal descriptions in the scoring guide. A rating should match the anchor paper's example. Also, when scoring all scorers should look first at a level three and then move up, down, or stay at a three.

Drift: This is the tendency over time for a reader to adjust a personal rating strategy up or down the scoring guide's scale because of the influence of the papers recently read. This results in errors and produces problems of low reliability. For example, after reading a lot of easy papers, future readings tend to be higher or after reading a series of good but hard papers the subsequent papers get lower scores because they don't match up to the caliber of the earlier papers.

The solution to this problem is to schedule regular intervals for your scoring and have a variety of papers to rate, not just students in the gifted or special education classes. Also, take and pass at regular intervals a "Drift" paper. This is a pre-scored paper from the same district. Rating it as the district has rated it helps a reader to come back to an agreed upon standard for scoring up and down the scoring guide's scale. Pre-scored "drift" papers readjust the personal "drift" back to the district's agreed to standard, i.e., as illustrated by anchor papers.

Fatigue and Boredom: Readers get tired. Reading several (hundreds of) essays on the same topic will make any one bored. Remember, each writer is attempting a unique response. He or she is writing just to you! So be prepared! Know when to say when! Breaks must be scheduled to interrupt the monotony of rating. Readers must take frequent breaks, and move around during the breaks.

Skimming: When readers get tired or bored, they tend to skim over a paper and judge it on how the first or last paragraph looks. This is always a mistake. Many papers improve significantly as the writing goes on, while others that start out strong cannot sustain the quality throughout. Every writer should receive equal attention by the reader to the quality of all of his or her writing work.

Suggestions

You are scoring a writing sample from a writing process that gave the student most likely two to four days from start to finish. Thus, (most likely) you won't see the quality of writing in these papers that you will find when students have time to engage thoughtfully in the entire writing process and get assistance from others with their drafts. Don't rate lower; just adjust your perspective knowing the writing environment.

Internalize the writing standards (anchor papers) agreed upon for scoring and apply only those -- regardless of your personal beliefs. Just as each child is different, so is each district's application of "standards."

It is the **preponderance of evidence** that determines which score you award, not one characteristic over another. Be mindful that learning doesn't progress all of its parts at the same time. Some skills move faster or slower than others. It is the overall movement you should use for your rating judgments. When rating, **assume the student starts with a middle score**. Then using the scoring guide in conjunction with the anchor papers adjust the score up or down from the middle. Use the image of a teeter-totter with strengths on one side and weaknesses/challenges on the other. Where is the preponderance of the evidence? Strengths or weaknesses? Up or down from the middle score? Fundamentally, the **middle score of the scoring guide** means the student can communicate. There may be problems and things still to learn but the message is understandable by the audience for which it is intended.

Having said three above, also be aware that in some districts **some traits or some criteria are more important** than others. These should be weighted more heavily in your judgment. Know your district. Learn in which traits and criteria this is true.

Always **read a paper at least twice**. The first read is to answer the question of who or what is in control of the writing process, the writer or the writing. Writers-in-control generally have scores in the middle or above on the scoring guide. When the writer is struggling with yet-to-be-in-control the scores are below the middle of the scoring guide's range.

Your second read of the paper is to award the trait scores.

Don't allow the repetition of similar errors by a student to influence your ratings unduly. The rule has simply not yet been learned by the student. It might be better to ask yourself the question of how much editing would be required to fix it. A lot of editing usually means lower levels of writing achievement. Simple editing might occur in the middle of the scoring guide range. Use the idea of editing to help award scores on any trait.

Following from above, **don't let your ratings on one trait influence your ratings on another trait**. It is true that some traits correlate with other traits, e.g., high on trait A often shows up as high on trait B.

The traits appear different (when we are teaching them) from lesson to lesson but in fact the traits may score very much alike. Sentence Fluency, for example, correlates highly with Conventions probably because sentences need correct punctuation to be effective as "pacers" or having a sense of rhythm (Sentence Fluency attribute). Yet Sentence Fluency is taught

differently from Conventions. Word Choice and Voice often correlate because words or phrases provide insight into the writer and what the writer wants her/his audience to feel/understand. Smaller vocabularies drive down the breadth that a writer's voice might take.

Having said all this our task is to NOT let one trait's mark in the sand affect where we place the mark for another trait. If traits are alike in scores, fine; if they are not alike then let the ratings from the scoring guide and anchor papers clearly show where the traits separate.

Boring is not a trait on which you can score! Writing in a testing situation is hard enough as it is. Don't expect a lot of flair but do respect the writing as a piece of that student's life.

Scoring in Conventions will vary slightly depending upon the prompt/topic. However, spelling and usage issues will be about the same regardless of the topic.

No matter what order you read and rate the traits, be aware you are **looking for consistency of your score within the trait across the whole paper (content across the whole paper); not consistency across the traits (a 3 in Content does not necessarily mean a 3 in Organization)**. The evidence you use to award a score must be evident throughout the paper, not just in one part of the paper.

If after all these suggestions you find yourself having a problem in scoring one trait or another, **seek assistance** from a team member. We are better all together than we are alone. Each of us has something to share. Each of us has a piece of the best answer. Together the picture is clearest.