

Norming Practices for Scoring Student Essays

Viktoria Gramp,
ELP instructor, Writing Curriculum Chair
Kansas State University
MIDTESOL 2018



Outline

- What: Definition
- Why: Rationale
- How: Process
 - Preparation
 - Execution
 - Post-norming analysis
- Things learned
- Questions



What is norming?

- “Norming, or calibration, is a process that brings a group of faculty raters together to decide how to assess student work in a consistent way, so that regardless of which rater assesses the work, the rating falls within a close range. The process is generally overseen by a facilitator,.”

Office of Assessment of Teaching and Learning (2016), *Quick Guide to Norming on Student Work for Program-level Assessment*. Washington State University
<https://atl.wsu.edu/documents/2015/03/rubrics-norming.pdf/>



What is norming for us?

- *5-8 sections of the same Writing course
 - 11-18 students in each
- 5-9 instructors
- Consistency is important => **3 norming sessions** a semester
- Scores should reflect true language proficiency & achieved SLOs
- Same sets of norming samples for 2-3 semesters

Why is norming important?

- Final essays are exchanged and double-graded =>
- Need to ensure inter-rater reliability
- Need to ensure consistency

- Benchmark essays

- Teachers adjust grading if needed



Daddy,
do you like
my picture?



Honey,
if you'd like me
to be objective,
I'll have to create
a rubric.



freshspectrum.com



Why is norming important?

Name: _____ Opinion Essay Rubric _____ 152/162/166 Writing _____ / 50 POINTS _____ %

10-9 = Excellent (Essay is coherent and fully-developed for the level; appropriate and sufficient details; may have minor lexical and grammatical errors)
8 = Good (Essay is mostly developed for the level; coherent but may display occasional redundancy and unclear logic; appropriate and sufficient details; occasional minor errors in structure, word form, and idiomatic language which do not interfere with meaning)
7 = Low Pass (Essay is generally effective and somewhat developed; occasional obscured ideas and mistakes in word form and idiomatic language which may result in lack of clarity; limited vocabulary and syntactic structures)
6-5 = Below level (Essay shows limited development; inadequate organization; inappropriate and insufficient support; multiple errors in sentence structure and word choice and form which impede meaning)
4-3 = Poor (Essay is seriously underdeveloped, disorganized, or not connected to the topic; little or no support; serious and frequent errors in sentence structure or usage that impede meaning)
2-1 = Not enough to evaluate (Essay is not connected to the topic; shows lack of understanding of the assignment)

*1/2 points are possible
**** a score of 7 in each category (70%) is the minimum standard for level-appropriate work**

| Opinion Essay Criteria | 10-9 | 8 | 7 | 6-5 | 4-3 | 2-1 |
|--|------|---|---|-----|-----|-----|
| Content and Development <input type="checkbox"/> Essay clearly addresses the prompt <input type="checkbox"/> Essay conveys opinion concisely <input type="checkbox"/> Fully developed thesis statement expressing an opinion <input type="checkbox"/> Clear and narrow focus <input type="checkbox"/> Smooth and interesting flow <input type="checkbox"/> Logical connection between ideas <input type="checkbox"/> Clear main ideas with sufficient support <input type="checkbox"/> Appropriate length for task | | | | | | |
| Organization <input type="checkbox"/> Introduction with hook, background or connecting information, and thesis statement <input type="checkbox"/> Focused paragraphs with clear topic sentences and convincing and specific supporting details <input type="checkbox"/> Cohesion and unity in paragraphs <input type="checkbox"/> Appropriate transition words and phrases show relationships between ideas | | | | | | |



How: Process

Prior to a norming session:

- **Scheduling:** T or U; 1h.15 min
- **Student work:** select a range– 6-7
- **Copied materials:** essay prompt, rubric, student work, score sheet, score sheet for recorder (**folders**)
- **Role:** Facilitator + timekeeper + recorder



How: Process

During a norming session:

- Rubric – questions?
- Distribute folders
- Allow time to read the task and the prompt(s)
- 7 minutes (silence)
- Record the scores



How: Process

During a norming session (continued):

Goal: discuss and come to consensus (close scores within a few points)

Rubric descriptions should be interpreted and applied in a similar way.



Examples

Let's look at a few examples



Scores: **Green** – too low; **purple** – too high

| Teachers | Your Score out of 100 (50 + 50) | | | | | | |
|---|---------------------------------|------------|------------------|--------------|-------------|-----------------|------------------|
| | #1 | #2 | #3 | #4 | #5 | #6 | #7 |
| Teacher 1 | 83 | 69.5 | 78.5 | 64 | 84 | 85 | 60 |
| Teacher 2 | 94 | 62 | 70 | 72 | 83 | 80 | 62 |
| Teacher 3 | 82 | 69 | 79 | 70 | 81 | 87 | ? |
| Teacher 4 | 86 | 61 | 77 | 61 | 76 | 79 | 61 |
| Teacher 5 | 81 | 68 | 85 | 64 | 79.5 | 85 | 59.5 |
| Teacher 6 | 83 | 70 | 73 | 64 | 81 | 87 | 57 |
| Teacher 7 | 86 | 64 | 88 | 67 | Not passing | 90 | 62 |
| Teacher 8 | 90 | 72 | 70 | 61 | 89 | 85 | 57 |
| Teacher 9 | 95 | ? | 73 | 64 | 79 | 81 | 62 |
| Viktoria 154-1 (3 hr) | 85 | 67 | 86 | 69 | 79 | 90 | 62 |
| Originally assigned scores Spring 2015 | 85 Clear pass | 68 fail | 86 Clear pass | 70 Border | 78 Pass | 93 High pass | 66 Clear fail |

1 point = 2% All 10s = 50/50 = 100% (A) All 9.5s = 47.5/50 = 95% (A) All 9s = 45/50 = 90% (A) All 8.5s = 42.5/50 = 85% (B) All 8s = 40/50 = 80% (B)

All 7.5s = 37.5/50 = 75% (C) A score of 7 in each category = 35/50 = 70% (low pass)



Summary & Response

Discussing Norming Samples (prior to CT 3)

Scores: **Green** – too low; **purple** – too high

| Teachers | Your Score out of 100 (50 + 50) | | | | | | | |
|---|---------------------------------|------------|------------------|--------------|----------------|-----------------|------------------|----|
| | #1 | #2 | #3 | #4 | #5 | #6 | #7 | |
| 152-1 | 80 | 67 | 87 | 71 | 69 | | | |
| 152-2 | 85 & even higher | 69 | 86 | 70 | 78 | | | |
| 152-3 | 95 | 73 | 86 | 77 | 84 | 90 | 64 | |
| 152-4 | 90 | 71 | 76 | 65 | 78 | | | |
| 152-5 | 84 | 63 | 75 | 67 | 68 | | | |
| 152-6 | 75 | 68 | 88 | 67 | 86 too high | 72 too low | 88 | 67 |
| 166-1 (3 hr) | 80 | 66 | 75 | 64 | 70 | 90 | 60 | |
| 150-7 FT | 94 | 70 | 82 | 69 | 86 | 88 | 57 | |
| 150-8 FT | 95 | 76 | 81 | 65 | 85 | 78 | 63 | |
| Xinran (GTA) | 89 | 67 | 66 | 76 | 71 | 91 | 63 | |
| Viktoria 154-1 (3 hr) | 85 | 67 | 86 | 69 | 79 | 90 | 62 | |
| Viktoria 162-1 (6 hr.) | 84 | 72 | 82.5 | 69 | 75 | 90 | 62 | |
| Originally assigned scores Spring 2015 | 85 Clear pass | 68 fail | 86 Clear pass | 70 Border | 78 Pass | 93 High pass | 66 Clear fail | |

1 point = 2%

All 10s = 50/50 = 100% (A)

All 9.5s = 47.5/50 = 95% (A)

All 9s = 45/50 = 90% (A)

All 8.5s = 42.5/50 = 85% (B)

All 8s = 40/50 = 80% (B)

All 7.5s = 37.5/50 = 75% (C)

A score of 7 in each category = 35/50 = 70% (low pass)



Comparison/Contrast Norming Samples (prior to Coordinated Test 2) **green:** the score is too low; **purple/pink:** the score is too high



| Teachers | Your Score out of 50 | | | | | |
|---------------------------|---|---|--|--|------------------------------------|---|
| | Essay #1 | Essay #2 | Essay #3 | Essay #4 | Essay #5 | Essay #6 |
| A | 5-5-5-5-5 = 25 (50%) | 6-6-7-7-7 = 33 (66%) | 8-8-7-6-7.5 = 36.5 (73%) | 9-9-8-7-9 = 42 (84%) | 9-9-9-9-9 = 45 (90%) | 10-10-8-9-10 = 47 (94%) |
| B | 4-4-5-4-4 =21 (42%) | 8-7-8-7-7=37 (74%) | 8-7-8-8-8=39 (78%) | 9-9-8-8-8=42 (84%) | 9-9-9-8-8=43 (86%) | 9-9-10-8-8=44 (88%) |
| C | 4-4-5-5-5 = 23 (46%) | 7-7-9-9-9 = 41 (82%) | 7-6-8-7-7 = 35 (70%) | 9-9-9-8-8 = 43 (86%) | 9-10-10-9-9=47 (94%) | 10-10-9-9-9 = 47 (94%) |
| D | 3-5-3-3-4=18 (36%) | 5-6-8-6-7 = 32 (64%) | 8-8-7.5-7-7 = 37.5 (75%) | 8.5-9-9-7.5-7.5 = 41.5 (83%) | 9.5-9-9.5-8.5- 9.5 = 46 (92%) | 10-10-8.5-8-10 = 46.5 (93%) |
| E | 5-6-6-6.5-6 = 29.5 (59%) | 7-7-7.5-7.5-7.5 = 36.5 (73%) | 7.5-8-7.5-7.5- 7.5-7.5 =38 (76%) | 8-8.5-8-8-8 = 40.5 (81%) | 8.5-8-8.5-8.5- 7.5 = 41 (82%) | 9-8.5-9-9-8 =43.5 (87%) |
| F | 4-5-5-5-4=23 (50%) | 6-6-6-6-6=30 (60%) | 7-7-7-7-7=35 (70%) | 8-8-7-7-7=37 (74%) | 9-8-8-7-7=39 (78%) | 10-8-8-8-8=42 (84%) |
| G | 2-4-4-4-4 = 18 (36%) | 4-5-6-5-5 = 25 (50%) | 6.5-6-7-6.5-6.5 =32 (65%) | 9-9-7-7-8 = 40 (80%) | 7.5-7-7.5-7-7 36 (72%) | 9-8-8.5-8.5-8.5 =42.5 (85%) |
| H | 2-3-2-3-3=12 (24%) | 6-6-5-6-6=29 (58%) | 8-9-7-6-8=38 (76%) | 8-8-7-6-7=36 (72%) | 9-9-8-7-8=41 (82%) | 10-10-8-8-9=45 (90%) |
| Viktoria | 4-4-4-4-5=21 (42%) | 5-5-6-5-7=28 (56%) | 8-8-7.5-7-7 = 37.5 (75%) | 9-9-7-6-8= 39 (78%) | 10-10-8-9-9.5 =46.5 (93%) | 10-10-8-9-9= 46 (92%) |
| Originally assigned score | 4-5-5-5-5=24 (48%) Diane Grammar score: 28/50 | 5-6-7-8-7 = 34 (66%) Agnes Grammar score: 34/50 | 8-8-8-7-7=37 (74%) Linh Grammar score: 36/50 | 9-9-8-8-9 = 43 (86%) Linh Grammar score: 46/50 | 9-9-8-9-8=43 (86%) Agnes | 9-(10+1)-8-8-8 45 (90%) Grammar score: 31/50 |
| "OUR VERDICT" | FAIL | FAIL | C (70-75%) | Low B (80-84% range) | High B – low A (86-94%) | High B – low A (85-94%) |



How: Process

Post-norming analysis:

- Charts after session:
 - Shared right away
 - Can be used the following semesters as reference
 - Benchmark essays can be selected after having been graded 18-20 times



Observations

- Good/readable copies are important
- If essays were handwritten, do not have them typed before distributing to raters
- Room: enough space for 5-9 teachers/ desks or tables
- 7 minutes per essay
- Open discussion of scores
- Original scores are in no way our target (can be off)



Thank you!

Questions?



Sources Consulted:

Office of Assessment of Teaching and Learning (2016), *Quick Guide to Norming on Student Work for Program-level Assessment*. Washington State University

<https://atl.wsu.edu/documents/2015/03/rubrics-norming.pdf/>

Rhode Island Department of Education (2013), *Writing Calibration Protocol*. Providence, RI.

<http://www.ride.ri.gov/InstructionAssessment/Literacy.aspx>

Rhode Island Department of Education & the National Center for the Improvement of Educational Assessment, Inc., *Calibration Protocol for Scoring Student Work*. Providence, RI.

<http://www.ride.ri.gov/Portals/0/Uploads/Documents/Teachers-and-Administrators-Excellent-Educators/Educator-Evaluation/Online-Modules/Calibration Protocol for Scoring Student Work.pdf>

Scoring Rubric Group Orientation and Calibration (2017), University of Hawaii Manoa.

<http://manoa.hawaii.edu/assessment/howto/rubrics.htm#p6>

