Create Grading Consistency Across Teaching Teams

Amy Pinkerton, MIDT
Senior Instructional
Designer, BSPH CTL

Lauren Dana, Instructional
Designer, BSPH CTL

1

Featuring Special Guests from BSPH

Michelle L. Bedoya, JD, LLM
PhD Student, Department of
Health Policy and
Management

Annelise Brochier, MPH
PhD Student,
Department of Health
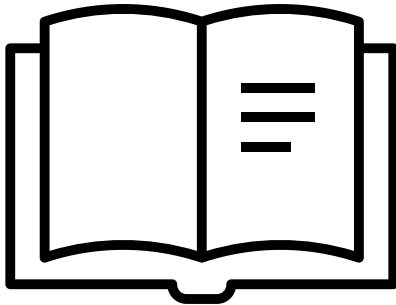Policy and Management

Anne Lilly, MSPH,
Senior Researcher,
Population, Family, and
Reproductive Health

Learning Objectives

Define Grade Norming

Explain Benefits and Process

Examine BSPH Case Studies

Definition, Benefits, and Process

"Grade Norming" Defined

Grade norming is a process that brings a **group** together to decide how to **assess** students' work **consistently** so that, regardless of which person grades the work, the rating falls within a close range.

Benefits for students

▶ Fairness and Equity

▶ Consistency in Feedback

▶ Increased Trust

(Office of Assessment for Curricular Effectiveness, 2020); (Johnson, n.d.)
Photo by Nathana Rebouças on Unsplash

Benefits for Teaching Teams

▶ Increased trust

▶ Encouragement of Collaborative Learning

▶ Save time

(Office of Assessment for Curricular Effectiveness, 2020); (Johnson, n.d.)
Photo by WOCInTech on Nappy.co

General Grade Norming Process

Select the Assessment

Create a Rubric or Standards

Model Grading

Practice Grading Individually

Compare and Calibrate Grading

Questions? 1

BSPH Case Studies

300.610.01 Public Health Policy

300.610.01 *Public Health Policy:* Context

Course Description

- ▶ Focuses on identifying and understanding decisionmakers, framing problems and key questions, developing and evaluating policy options, and advocating for change

- ▶ Students:
  - ○ 326 incoming MPH students

- ▶ Teaching team:
  - ○ Dr. Josh Sharfstein (Professor)
  - ○ 2 Lead TAs
  - ○ 10 TAs

300.610.01 *Public Health Policy*: Rationale

Why did we choose to implement grade norming?

► Level set with large teaching team

► Build self-confidence among TAs, many of whom were TAing for the first time

► Respond to feedback from prior years' course evaluations
  ○ Concerns about inconsistent grading
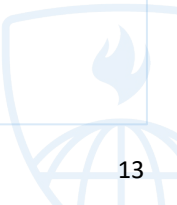
Assessing a Policymaker Assignment: Total Possible Score- 30 points

Task	Level 3: Excellent	Level 2: Satisfactory	Level 1: Poor	Total Points
Identification and description of policymaker (6 points)	Policymaker identified correctly with thorough summary of professional background and current position in government (6-5 points)	Policymaker identified, but summary of professional background and position in government lacks detail or has minor inaccuracies (4-3 points)	Policymaker identified incorrectly and/or summary of professional background and position in government lacks key information or has major inaccuracies (2-0 points)	
Description of policy (6 points)	Identifies topic as well as clearly describes a single, specific policy that the policymaker has sought to advance. (6-5 points)	Identifies public health topic of interest, but specific policy is not that specific or is poorly related. (4-3 points)	Conflates topic and policy or very few details. (2-0 points)	
Assessment of constraints on addressing public health issue (6 points)	Both formal and informal constraints on policymaker's ability to advance the specific policy discussed clearly (6-5 points)	Vague discussion of constraints, but both formal and informal are mentioned. (4-3 points)	Poor discussion of constraints, or no discussion. (2-0 points)	
Assessment of effectiveness in face of constraints (6 points)	Thoughtful assessment of policymaker's effectiveness given the formal and informal constraints (6-5 points)	Assessment of policymaker's effectiveness lacks detail, with little support from discussion of authority and constraints (4-3 points)	Assessment of policymaker's effectiveness lacks key information and/or is not aligned with authority and constraints (2-0 points)	
Writing quality (3 points)	Information is presented clearly and is easy to follow (3 points)	The clarity of the writing could be enhanced (2 points)	Grammatical and typographical errors make it challenging to read (1-0 points)	
Overall organization and length (3 points)	Information is presented in a clear and logical manner, and meets page number requirements (3 points)	Most of the assessment is organized in a clear way and/or does not meet page number requirements (2 points)	Information is presented in a way that makes it difficult to follow and/or does not meet page number requirements (1-0 points)	
Total Points				/30

Public Health Policy: SPH 300.610.01

Assessing a Health Policymaker

Instructions:

Find and assess a policymaker, in 6 steps. Please label the items in your memo by the numbers 1-6.
1. Identify an individual by name and title, with a photo. Briefly summarize their professional background, up to the point where they moved into this particular position.
2. Correctly classify the **individual's branch of government, type of position, and level of position**.
3. Describe **one** public health topic of interest to this person, summarizing their work in this area.
4. In the area of this public health topic (from #3 above), describe one example of **a specific policy** that the policymaker has sought to push forward.
5. Describe and discuss the constraints, both informal and formal, on this policymaker in advancing this **specific policy**.
6. Give your own opinion of how effective the policymaker is in addressing these constraints and advancing the policy, and provide reasons to explain your opinion.

Other information:

- Before starting this assignment, please review materials you have received on academic ethics, including plagiarism.
- Do not use ChatGPT or any other artificial intelligence source on the internet.
- You should not choose a policymaker we have interviewed in class.
- The paper should be formatted to be:
  - No more than 3 pages – **excluding references**
  - Single-spaced
  - 12-point Times New Roman font
  - 1-inch margins
  - Please reference any ideas, quotes, or paraphrased material. **If in doubt, include a citation.** Please cite references using endnotes (use in-text numbered citations and list all references at the end of the paper using any citation style you want – APA is common). References will not count against the 3-page limit

Q & A:

Who can I pick as a policymaker?
  You can pick any public sector policymaker for this assignment, as long as they have not been interviewed in class. The policymaker can come from any country. We recommend that you choose a policymaker who seems interesting to you! You should consider state (or regional) and local legislators, elected officials, health officials, judges, and others.

Where can I find information?
  We recommend using a variety of sources to find information about your policymaker, their interests, and authority. These include government websites, policymakers' personal websites or biographies, and news media sources among others. Regular google searches and the databases
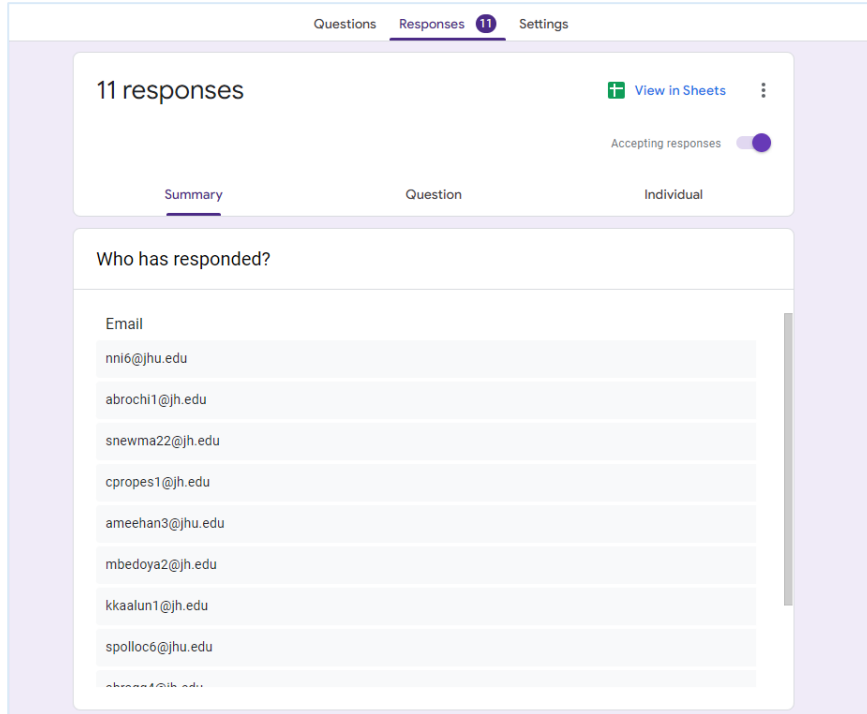
Our Process

▶ Introduce activity during first TA meeting
  ○ Reintroduce activity immediately before first assignment deadline

▶ Convert rubric into Google Form to facilitate "analysis" of grading decisions

▶ Share Google Form along with two student submissions with the TAs

▶ Disseminate "results" and key takeways via email

▶ Conduct norming asynchronously to accommodate scheduling constraints

Example 1 (6 point categories)

	Excellent (6 points)	Excellent (5 points)	Satisfactory (4 points)	Satisfactory (3 points)	Poor (2 points)	Poor (1 point)	Poor (0 points)
Identification and description of policymaker	☐	☐	☐	☐	☐	☐	☐
Description of policy	☐	☐	☐	☐	☐	☐	☐
Assessment of constraints on addressing public health issue	☐	☐	☐	☐	☐	☐	☐
Assessment of effectiveness in face of constraints	☐	☐	☐	☐	☐	☐	☐

300.610.01 *Public Health Policy*: Responses

► Two randomly selected assignment examples

► 4 six-point categories, 2 three-point categories adapted from the rubric

► Google Form responses
  ► 11 responses (12 TAs) = 91.67% TA response rate
  ► 6 domains × 11 TAs × 2 assignments = 132 observations for norming exercise

300.610.01 *Public Health Policy*: Data Visualization

Example 1 (6 point categories)
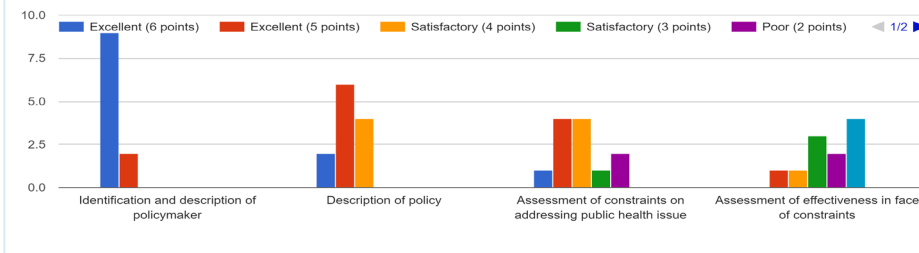
Example 2 (6 point categories)
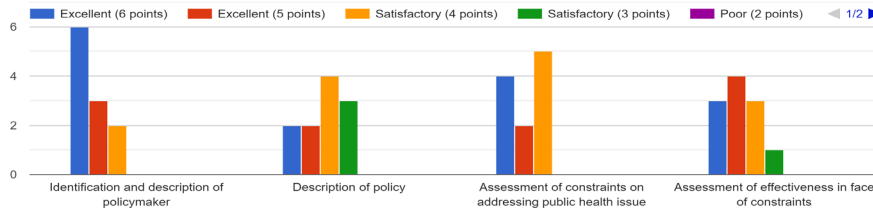
► Two ways to view Google Form responses
  ► Bar graphs
    ● Helpful to visualize variability in scores across the different categories of rubrics
    ● Allows to quickly see patterns, such as which categories have higher or lower scores and how scores differ across categories

17

► Two ways to view Google Form responses
  ► Spreadsheet data (more detailed analysis)
    ● Viewing raw scores in a spreadsheet
    ● Source for descriptive statistics like the mean or median score for each domain to check central tendencies and spread of scores.
    ● By analyzing trends and variations among TAs, helped identify any tendencies of certain TAs to grade more strictly or leniently, which helps in assessing grading consistency.

300.610.01 *Public Health Policy*: Steps

► **Step 1:** Collected and analyzed the raw scores from TAs and visualized the data using bar graphs to explore variability and trends across rubric categories for two sample assignments
  ► The bar graphs and raw data highlighted the need for further calibration to ensure alignment in grading expectations among TAs

► **Step 2:** Conducted exploratory analysis to identify key insights.
  ► This included calculating the mean scores for each category across both assignments to assess variability, and identifying categories with greatest variability
  ► We evaluated the data not only to pinpoint where scoring differences occurred, but also to understand the potential reasons behind these discrepancies
  ► This process involved a qualitative assessment to understand not just where scores differed, but also *why* these differences might have occurred
  ► Provided feedback and specific recommendations through email with the aim to standardize the grading process and promote consistent application of the rubric across all TAs

Team,

Thank you for completing our asynchronous norming exercise by scoring the attached Example 1 (U.S. policymaker) and Example 2 (international policymaker).

Here is a summary of the results from the norming exercise for the two example assignments. As you know, the scores are out of 6 for most categories and out of 3 for writing quality and overall organiz

Example 1
1. Identification and Description of Policymaker (6-point Category):
   o Mean Score: 5.82
2. Description of Policy (6-point Category):
   o Mean Score: 4.91
3. Assessment of Constraints on Addressing Public Health Issue (6-point Category):
   o Mean Score: 4.18
4. Assessment of Effectiveness in Face of Constraints (6-point Category):
   o Mean Score: 2.36
5. Writing Quality (3-point Category) :
   a. Mean Score: 2.18
6. Overall Organization and Length (3-point Category) :
   a. Mean Score: 2.60

Example 2
1. Identification and Description of Policymaker (6-point Category):
   a. Mean Score: 5.36
2. Description of Policy (6-point Category):
   a. Mean Score: 4.27
3. Assessment of Constraints on Addressing Public Health Issue (6-point Category):
   o Mean Score: 4.91
4. Assessment of Effectiveness in Face of Constraints (6-point Category) :
   o Mean Score: 4.82
5. Writing Quality (3-point Category):
   o Mean Score: 2.36
6. Overall Organization and Length (3-point Category):
   o Mean Score: 2.55

Bar graphs representing the scores for each example are attached. The bar graphs illustrate the variability in scoring among TAs for Examples 1 and 2.

In Example 1, the bars show significant variability in scoring "Assessment of Constraints on Addressing Public Health Issue" and "Assessment of Effectiveness in Face of Constraints," which indicates a lack

In Example 2, the bars show notable variability in scoring "Description of Policy," with some TAs scoring more critically than others, suggesting differences in expectations or interpretation. Similar to Exa

Altogether, the variability in the bar graphs indicates the need for further calibration and standardization in grading to ensure a more consistent assessment approach. To do so, please meticulously adhe maintain consistency. Please also:

- Use the full range of scores (6-0) to differentiate between excellent, satisfactory, and poor descriptions of a policymaker.
- Be cautious of conflating a topic with a policy. Ensure that the student specifies a policy that the policymaker has sought to advance, and assess the level of detail provided.
- Ensure that all aspects of constraints are considered and differentiate between a vague mention and a detailed explanation. A more detailed discussion should receive higher scoring.
- Be consistent in how you judge the depth and support of the analysis. A well-supported assessment should include specific examples and evidence of the policymaker's actions and their effective
- Apply the writing quality criteria evenly, considering both major and minor grammatical issues.
- Be consistent in applying the criteria for organization and length. If an assessment is hard to follow or deviates significantly from the required length, it should be marked down accordingly.
- Document your rationale for scores, especially for borderline cases, to ensure that your grading can be reviewed and understood by others.
- **MOST IMPORTANT**: Be aware of potential biases, such as familiarity with the topic or student, and strive to apply the rubric criteria objectively and consistently.
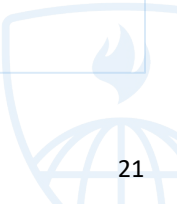
300.610.01 Key Takeaways - 1

- ▶ Positive outcomes
  - ▶ Enhanced teaching team confidence by clarifying expectations and identifying common grading tendencies
  - ▶ Identified inconsistencies and provided targeted guidance by highlighting areas with significant variability in scores to help TAs better understand how to apply the rubric consistently

- ▶ Contextual factors
  - ▶ TA engagement considerations
  - ▶ Assignment-specific challenges
    - We noted that students had varying interpretations of assignment instructions
    - Highlighted the need for clearer instructions and additional guidance

- ▶ Insights for improvement
  - ▶ Rubric and instruction refinement
    - Offered valuable insights into how both the rubric and assignment instructions could be refined

300.610.01 Key Takeaways - 2

► Norming is an ongoing process, not a one-time activity.

► While ideally norming would occur for every assignment, this may not be feasible due to the time constraints of TAs.
  o Begin by conducting a norming activity to set a common standard for grading expectations.
  o Establish a mechanism to maintain consistency in grading across the teaching team.
    ▪ For example, utilize annotated rubrics with clear grading guidelines to ensure uniformity.
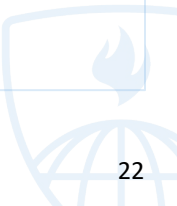
380.611.81 Fundamentals of Program Evaluation

Grade Norming in *380.611.81 Fundamentals of Program Evaluation*: Context

► Context of our class
- First term
- Online course, mostly asynchronous
- 80 to 100 students
- 3 to 4 TAs
- 4 written assignments

► Grade norming practices
- TA selection- performance task
- Grading Guides
- Teaching Team Rubric- Independent grading and then meeting to come to agreement on scores
- Grading Reliability Worksheet- track rubric scores during grading
- Questions- all grading questions posed to entire teaching team (WhatsApp or email)
- Review rubrics

Target population: • A primary and secondary target populations are described and correctly identified	5=primary and secondary populations are correctly identified and with specific characteristics	4=while primary and/or secondary populations are correctly identified, is missing key characteristic	3=incorrectly identifies either primary and/or secondary target populations	2=incorrectly identifies both primary and secondary target population	0=did not list target populations	**Target Population:** - Female adolescent orphans in Lusaka Zambia aged 14-18 years. - Secondary target populations could be female teachers - Note: target population is who you are trying to act on; the secondary population is the group who "could" see change but isn't the main focus of the intervention. **Grading this Aspect:** To receive 5 points, students must identify all of the aspects above. If one aspect is missing, give a score of 4 here. If the primary or secondary population is incorrectly identified, give a score of 3 here.
Activities: • Activities are mentioned that address the program objectives	10=all activities are correctly listed	9=missing one of the activities described or lists activity outside of program description	8=missing 2-3 of the activities or is listing an activity not described	7-1=did not list most of the activities described or includes multiple activities from outside of the description	0=activities are not listed	**Activities:** - Participating orphan girls were given money for school fees and supplies - Female teachers at each school were trained as helpers, who ensured participating orphan girls attended school - Fund at each school was established to assist helpers address attendance issues **Grading this Aspect:** - If students are missing one of the activities, take off 1 point in this area. - If missing 2-3 activities, consider give 8 points here. - If students provide more specificity in their activities based on the project description, that is fine. No need to take off points. - If students include evaluation activities (data collection for the evaluation, etc.), take off 1 point here if there are no other issues.

25

380.611.81 Tools: Teaching Team Rubric

Assignment 1: Description of project, creation of logic model, selection of indicators

Name of Student:

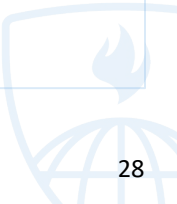| Component	Excellent	Good	Satisfactory	Unsatisfactory	Not Completed	AL	DS	KM	MM	ZL	Agreed Upon Score
Goal: • A broad goal is mentioned and phrased appropriately	5= has all criteria of goal	4=missing one component (might not have listed population or geography or impact)	3=missing two components, such as population, geography and/or impact	2=missing or incorrectly identified several components (population, geography and/or impact) and/or is not written as a broad statement	0=no goal written **AL Notes: Nicely done- clear and includes the important info!**	5	5	5	5	5	5
Objectives: • Follows SMART criteria; at population level	10=each objective correctly has the SMART criteria and is at the population level	9=missing one of the SMART criteria and/or one objective not at population level	8=missing 2-3 of the SMART criteria in at least one objective and/or 2-3 objectives are not at population level	7-1 =missing most of the SMART criteria in at least two objectives; and/or none of the objectives are at population level	0=no objectives written **AL Notes: Overall, very good job demonstrating SMART criteria! Missing the 14-18 in objective #1. And missing Lusaka in third objective. [Note about pre-test/post-test.]**	9	10	10	9	10	9

380.611.81 Tools: Grading Reliability Worksheet

Group	Max Points	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	Nigeria Average	Nutrition Average	Zambia Average
Goal	5	4	5	4	5	4	4	4	4	4	5	4	5	5	4	4	5	4	5	4	5	5	5	5	5	5	4	5	5	5	4.2	4.5	4.9
Objectives	10	8	9	9	7	10	7	8	10	8	7	4	10	9	10	10	9	7	7	10	9	8	9	8	8	9	9	10	7	8	8.4	8.3	8.5
Need	5	5	5	5	2	4	5	5	3	5	5	5	5	3	5	4	5	4	4	5	5	5	5	5	5	3	5	5	3	5	4.3	4.5	4.6
Target Population	5	5	5	5	5	5	5	5	4	5	5	5	2	5	3	2	5	4	5	4	5	5	5	5	5	4	5	5	5	5	4.9	4.0	4.9
Activities	10	10	9	9	10	10	9	9	10	10	10	10	10	10	8	9	10	7	9	10	8	10	10	10	9	10	9	10	10	10	9.6	9.3	9.6
Logic Model	30	27	24	27	24	24	25	21	24	25	29	26	28	27	26	24	29	26	26	26	25	26	26	25	26	26	28	28	24	24	24.6	26.7	25.8
Indicators- #	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10.0	10.0	10.0
Input Indicators	5	5	5	4	4	5	4	5	4	5	5	5	5	5	4	5	5	5	4	4	4	5	4	5	3	3	5	5	5	4	4.6	4.7	4.3
Activity Indicators	5	4	4	4	5	5	4	5	4	5	5	5	4	5	5	5	2	5	4	5	4	4	4	4	5	5	5	4	4	3	4.4	4.5	4.2
Output Indicators	5	5	5	4	4	4	4	5	3	5	5	5	4	4	4	5	4	5	3	4	4	4	4	5	4	5	3	4	5	4	4.3	4.3	4.2
Outcome Indicators	5	3	3	3	3	5	2	3	4	2	5	2	2	5	4	5	2	4	5	3	5	3	4	2	3	4	3	3	3	4	3.1	3.7	3.4
Impact Indicators	5	4	4	4	4	5	5	4	5	5	5	5	1	4	4	5	1	4	4	3	4	4	5	2	5	5	4	4	3	3	4.4	3.6	3.9
Late Submission?																																	
TOTAL SCORE	100	90	88	88	83	91	84	84	85	89	96	86	86	92	87	88	87	85	86	88	88	89	91	86	88	90	89	93	84	85	86.9	88.1	88.3

27

380.611.81 Key Takeaways

▶ Getting reliable on grading takes a lot of work!

▶ Good to employ multiple strategies

▶ OneDrive is a great tool
  - Multiple teaching team members can update Teaching Team Rubric and Grading Reliability Worksheet at the same time without conflicted copies.

▶ Open communication is essential!
  - Important to have means through which TAs know how they are grading compared to others
  - Important to have strategy for getting grading questions answered quickly and for answers to be shared with whole team.

Questions? 2

Practice!

Instructions - 1

▶ Step 1: Access the Grade Norming Workshop Worksheet.

▶ Step 2: Complete each of the steps for the Sample Grade Norming Activity.
- Select the Assignment
- Create (Use) a Rubric
- Model Grading

Model Grading for Sample 1

Mental health is a problem in my community. A lot of people have anxiety and depression. It's becoming more common, and some people can't get help because they can't afford it.

► **Impact:** Different groups in the community are affected, but people who don't have much money seem to struggle more. Also, young people have a lot of stress.

► **Interventions:** Giving more people access to therapy and maybe having some mental health lessons in schools could help. These things might make it easier for people to feel better and get the help they need (Open AI, 2024).

► *No references included.*
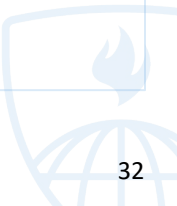
Model Grading for Sample 2

Mental health disorders, including anxiety and depression, are on the rise in my community, driven in part by economic stress, social isolation, and the lingering effects of the COVID-19 pandemic. The demand for mental health services has outpaced available resources, particularly for vulnerable groups.

► **Impact:** Low-income individuals, racial minorities, and young adults are disproportionately affected due to barriers like stigma, lack of insurance, and limited access to culturally competent care. These populations experience higher rates of untreated mental health issues, leading to increased risk of homelessness, substance abuse, and suicide.

► **Interventions:** Expanding telehealth mental health services can provide more accessible care, particularly for those in underserved areas. Additionally, integrating mental health services into primary care settings and increasing mental health education and outreach in schools can help reduce stigma and provide early intervention (Open AI,2024).

► *3-4 References from course materials included.*
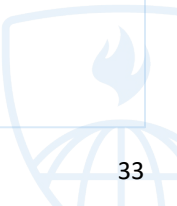
Instructions - 2

▶ Step 1: Access the Grade Norming Workshop Worksheet.

▶ Step 2: Complete each of the steps for the Sample Grade Norming Activity.
  • Select the Assignment
  • Create (Use) a Rubric
  • Model Grading
  • Practice Grading Individually
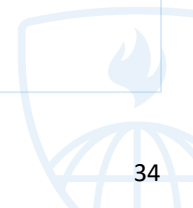
Practice and Compare Sample 3

Climate change is becoming a concern in my community, with increasing temperatures and unpredictable weather patterns. These changes are affecting public health through rising rates of heat-related illnesses and respiratory issues due to poor air quality.

► **Impact:** Vulnerable populations, such as the elderly, children, and low-income families, are disproportionately impacted. These groups may lack access to proper cooling systems or live in areas with higher pollution levels, worsening their health outcomes.

► **Interventions:** Increasing public awareness through community education programs and investing in green infrastructure, like parks and tree planting, could help reduce the effects of climate change and promote healthier environments (OpenAI, 2024).

► *1 reference from course materials included.*

Practice and Compare Sample 4

In Baltimore, climate change is intensifying the urban heat island effect, causing higher temperatures and more frequent heatwaves. Rising sea levels and increased flooding also threaten vulnerable neighborhoods, particularly those along the waterfront. These environmental shifts are contributing to a range of public health issues, including heat-related illnesses and worsening asthma rates due to air pollution.

► **Impact:** Low-income communities and communities of color in Baltimore are disproportionately affected. These populations often reside in areas with fewer green spaces and older housing, making them more vulnerable to extreme heat, flooding, and poor air quality.

► **Interventions:** Implementing green infrastructure initiatives, such as expanding tree canopy coverage and creating more parks, could reduce the urban heat island effect and improve air quality. Additionally, investing in flood resilience programs and improving stormwater management systems can protect vulnerable neighborhoods from rising sea levels and heavy rainfall events (OpenAI, 2024).

► *2 references from course materials included.*
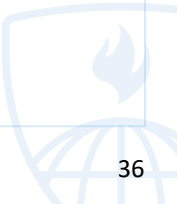
Practice and Compare Sample 5

Climate change is a big issue in California. The state has been experiencing more wildfires and hotter temperatures in recent years. This has caused some health problems for people living in affected areas.

► **Impact:** Some people, like those who live in areas where wildfires happen, are affected more than others. Also, people with breathing problems find it hard when there is smoke in the air.

► **Interventions:** Planting more trees and using less water could help. More people should also be taught about how climate change is affecting the state so they can take action (OpenAI, 2024).
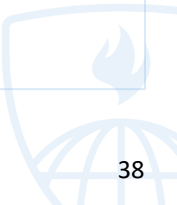
► *2 references from course materials included.*
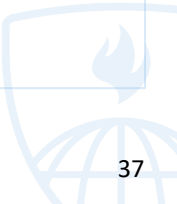
Instructions

▶ Step 1: Access the Grade Norming Workshop Worksheet.

▶ Step 2: Complete each of the steps for the Sample Grade Norming Activity.
- Select the Assignment
- Create (Use) a Rubric
- Model Grading
- Practice Grading Individually
- Compare and Calibrate

▶ Step 3: Repeat the process with one of your course assignments.

How did we do? 2-minute anonymous workshop evaluation

Scan the code or go to:
**https://forms.office.com/r/
Gs1pzkrgSQ**

**BSPH CTL Teaching Toolkit
Workshop Evaluation AY24-25**

Questions? 3

References

Johnson, J. (n.d.). *Consistency and Equity in Grading*. The Derek Bok Center for Teaching and Learning.
https://bokcenter.harvard.edu/consistency-equity-grading

Office of Assessment for Curricular Effectiveness. (2020, December 18). *Quick Guide to Norming on Student Work for Program-level Assessment*. Washington State University.
https://ace.wsu.edu/documents/2015/03/rubrics-norming.pdf/

Teaching with Writing. (n.d.). *Running a Grade-Norming Session*. University of Minnesota.
https://wac.umn.edu/tww-program/teaching-resources/grade-norming-session
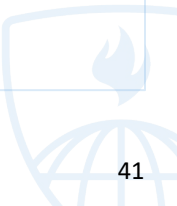
Resources

- Past CTL Workshops:
  - *Giving Students Effective Feedback* (February 2024)
  - *Rubrics: Benefits and Grading Strategies* (September 2023)

- CTL's Teaching Toolkit *Assessment and Grading* resource page

- *Grade Norming Activity* worksheet that accompanies this workshop