

Performance Assessment Review

Trial Results & Interpretation



COLORADO DEPARTMENT OF EDUCATION

COLORADO COMMUNITY COLLEGE SYSTEM



Overview



- **Trial Data**

- What information can we use to review the tasks and rubrics?
- What were some of the common strengths/issues that emerged from the trials?

Task Quality: Cognitive Rigor



- **Examine the cognitive rigor of your tasks**
 - Hess Cognitive Rigor Matrix
- **Performance tasks should have components that are written at DOK Level 3 or DOK Level 4**
 - Ask yourself—are the rubric criteria written at the appropriate level of complexity?

Task Quality: Cognitive Rigor



Hess Cognitive Rigor Matrix & Curricular Examples: Applying Webb's Depth-of-Knowledge Levels to Bloom's Cognitive Process Dimensions

Revised Bloom's Taxonomy	Webb's DOK Level 1 Recall & Reproduction	Webb's DOK Level 2 Skills & Concepts	Webb's DOK Level 3 Strategic Thinking/ Reasoning	Webb's DOK Level 4 Extended Thinking
Remember Retrieve knowledge from long-term memory, recognize, recall, locate, identify	<ul style="list-style-type: none"> Complete short answer questions with facts, details, terms, principles, etc. (e.g., label parts of diagram) 	Use these Hess CRM descriptors with most writing and oral communication assignments or assessments in any content area.		
Understand Construct meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarize, generalize, infer a logical conclusion, predict, compare/contrast, match like ideas, explain, construct models	<ul style="list-style-type: none"> Describe or define facts, details, terms, principles, etc. Select appropriate word/phrase to use when intended meaning/definition is clearly evident Write simple complete sentences Add an appropriate caption to a photo or illustration Write "fact statements" on a topic (e.g., spiders build webs) 	<ul style="list-style-type: none"> <u>Specify, explain, show relationships</u>: explain why, cause-effect Provide and explain non-examples and examples Take notes; organize ideas/data (e.g., relevance, trends, perspectives) <u>Summarize</u> results, key concepts, ideas Explain central ideas or accurate generalizations of texts or topics Describe steps in a process (e.g., science procedure, how to and why control variables) 	<ul style="list-style-type: none"> Write a multi-paragraph composition for specific purpose, focus, voice, tone, & audience Develop and explain opposing perspectives or connect ideas, principles, or concepts using supporting evidence (quote, example, text reference, etc.) Develop <u>arguments of fact</u> (e.g., Are these criticisms supported by the historical facts? Is this claim or equation true?) 	<ul style="list-style-type: none"> Use multiple sources to elaborate on how concepts or ideas specifically draw from <i>other</i> content domains or differing concepts (e.g., research paper, <u>arguments of policy</u> – should this law be passed? What will be the impact of this change?) Develop generalizations about the results obtained or strategies used and apply them to a new problem or contextual scenario

© 2009 (updated 2013) Karin K. Hess. Hess Cognitive Rigor Matrix for Writing: Applying Argument Types Across Content Areas to the CRM. Copy/use only with full citation
Khes@nciea.org

Task Quality: Cognitive Rigor



Webb's DOK Level 2
Skills & Concepts

Webb's DOK Level 3
Strategic Thinking/ Reasoning

Use these Hess CRM descriptors with most w
communication assignments or assessments in

- Specify, explain, show relationships; explain why, cause-effect
- Provide and explain non-examples and examples
- Take notes; organize ideas/data (e.g., relevance, trends, perspectives)
- Summarize results, key concepts, ideas
- Explain central ideas or accurate generalizations of texts or topics
- Describe steps in a process (e.g., science procedure, how to and why control variables)

- Write a multi-paragraph composition for specific purpose, focus, voice, tone, & audience
- Develop and explain opposing perspectives or connect ideas, principles, or concepts using supporting evidence (quote, example, text reference, etc.)
- Develop arguments of fact (e.g., Are these criticisms supported by the historical facts? Is this claim or equation true?)

Task Quality: Cognitive Rigor



Webb's DOK Level 2 Skills & Concepts

Webb's DOK Level 3 Strategic Thinking/ Reasoning

Use these Hess CRM descriptors with most w
communication assignments or assessments in

- Specify, explain, show relationships; explain why, cause-effect
- Provide and explain non-examples and examples
- Take notes; organize ideas/data (e.g., relevance, trends, perspectives)
- Summarize results, key concepts, ideas
- Explain central ideas or accurate generalizations of texts or topics
- Describe steps in a process (e.g., science procedure, how to and why control variables)

- Write a multi-paragraph composition for specific purpose, focus, voice, tone, & audience
- Develop and explain opposing perspectives or connect ideas, principles, or concepts using supporting evidence (quote, example, text reference, etc.)
- Develop arguments of fact (e.g., Are these criticisms supported by the historical facts? Is this claim or equation true?)

Task Difficulty



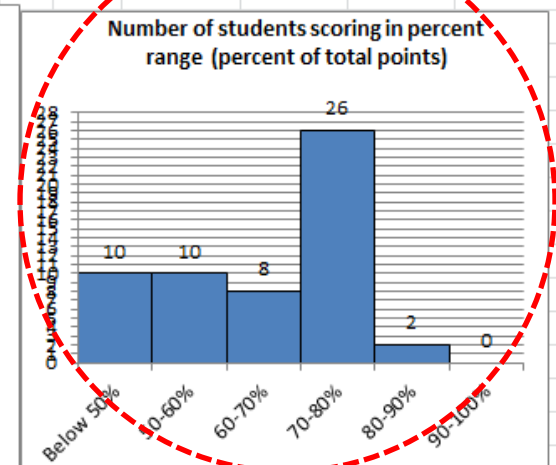
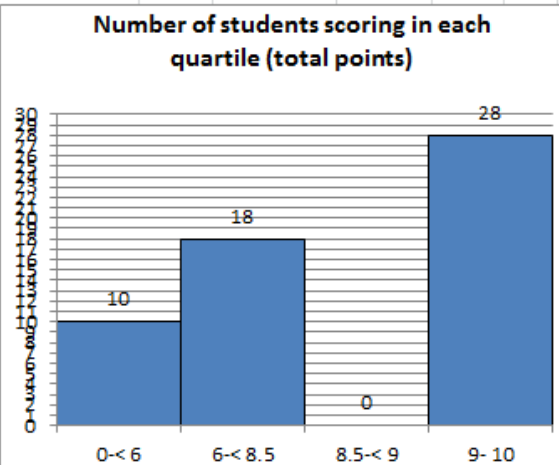
- **Think about:**
 - What level of performance did you expect to see with your students?
 - How did students actually perform on the task?
- **Refer to information in the:**
 - Data reports
 - Comments that teachers provided during the trials

Task Difficulty (Overall)



“Report-main” Tab

	A	B	C	D	E	F	G	H	I	J	K	L
1	Content Area		Health									
2	Grade Level		8th Grade									
3	Task		Media Critique and Analysis									
4	Criteria	Description	N	Mean	SD	Min	Max	"0" Scores				
5	1		56	3.07	0.85	0	4	2				
6	2		56	2.00	1.25	0	3	13				
7	3		56	2.30	1.25	0	3	13				
8	4		0									
9	5		0	#####								
10	6		0									
11	7		0									
12	8		0									
13	9		0									
14	10		0									



Task Difficulty (Overall)



“Weight-pt-report” Tab

- Column A “Percent of total points”
- Column C “Total Points/Student” (raw scores)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1				Please input a weight for each Criteria. The report will adjust										
2				based on the weight you enter										
3				Weight x#	1	3	1							
4				max points	4	12	4							
5				Criteria	1	2	3	4	5	6	7	8	9	10
6	Percent of total points	Weighted Average	Total Points / student	Student Number / Identifier										
7	15.0%	0.6	3	s1	3	0	0							
8	65.0%	2.6	13	s2	4	6	3							
9	30.0%	1.2	6	s3	0	3	3							
10	55.0%	2.2	11	s4	3	6	2							
11	65.0%	2.6	13	s5	2	9	2							

Task Difficulty (Criterion Level)



“Report-main” Tab

- Mean
 - ✦ Histogram
- SD (spread of scores)
- Min/Max

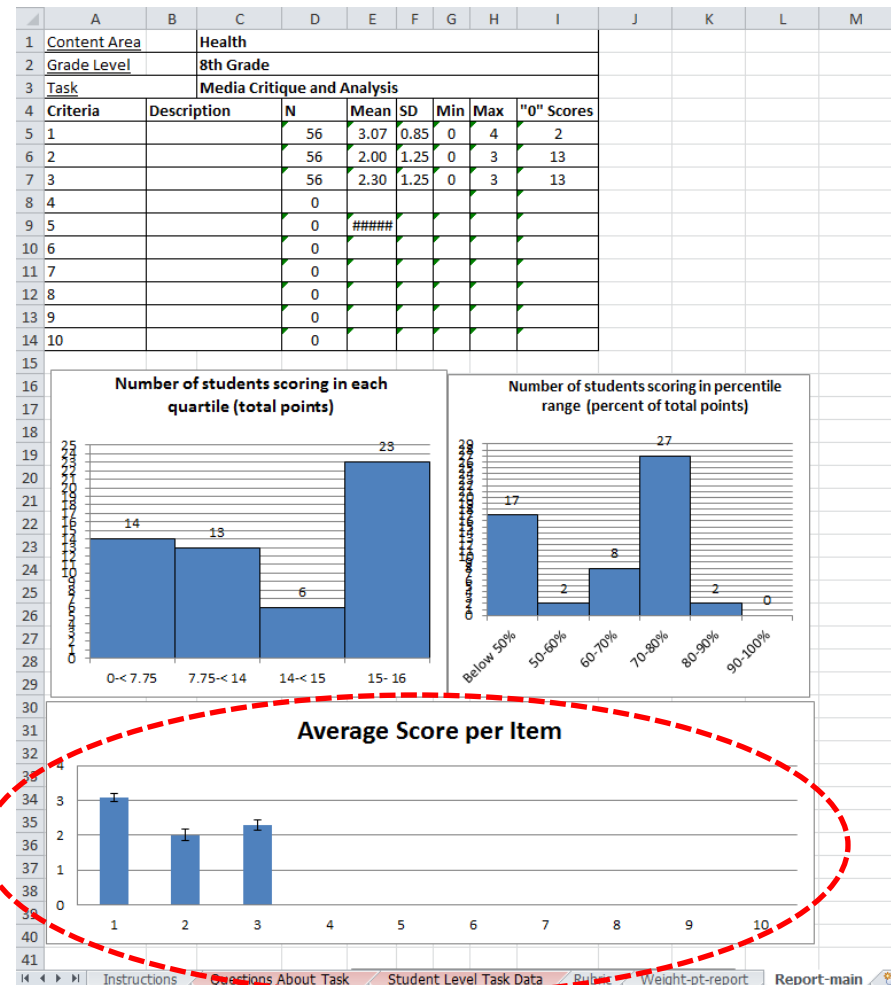
	A	B	C	D	E	F	G	H	I
1	Content Area		Health						
2	Grade Level		8th Grade						
3	Task		Media Critique and Analysis						
4	Criteria	Description	N	Mean	SD	Min	Max	"0" Scores	
5	1		56	3.07	0.85	0	4	2	
6	2		56	2.00	1.25	0	3	13	
7	3		56	2.30	1.25	0	3	13	
8	4		0						
9	5		0	#####					
10	6		0						
11	7		0						
12	8		0						
13	9		0						
14	10		0						

Task Difficulty (Criterion Level)

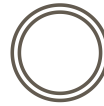
“Report-main” Tab

○ Average Score per Item”

✦ Note: *Item = Criterion*



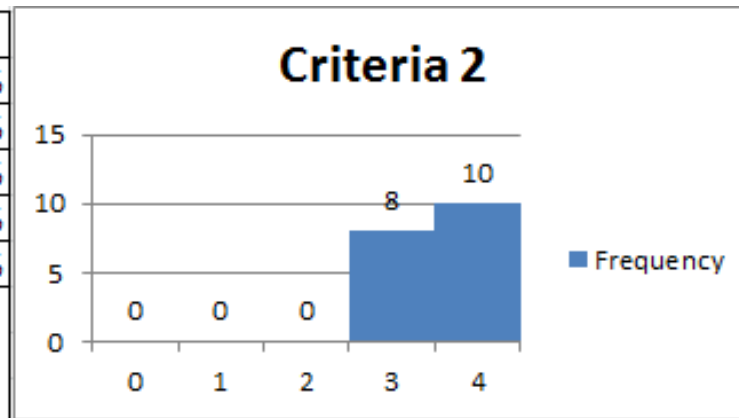
Task Difficulty (Criterion Level)



“Report-main” Tab—Histograms

- For each criterion, you can see:
 - Difficulty level
 - Distribution of scores
 - Areas in which there might be potential issues with the rubric

55	Criteria 2	Score	Frequency	Percent
56		0	0	0.0%
57		1	0	0.0%
58		2	0	0.0%
59		3	8	44.4%
60		4	10	55.6%
61	Comments:			
62				
63				

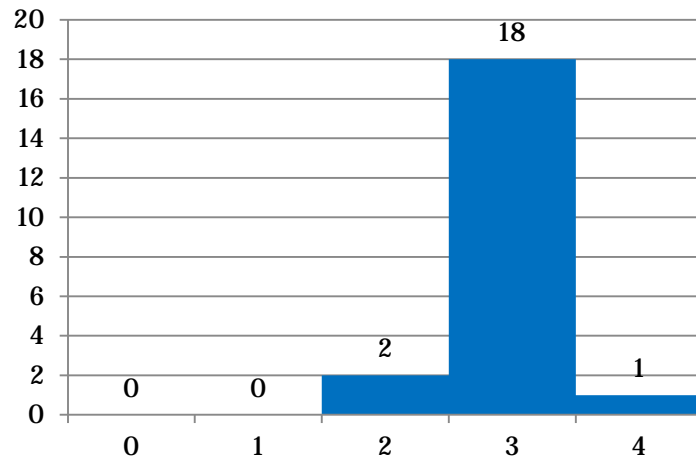


Criterion-Level Data

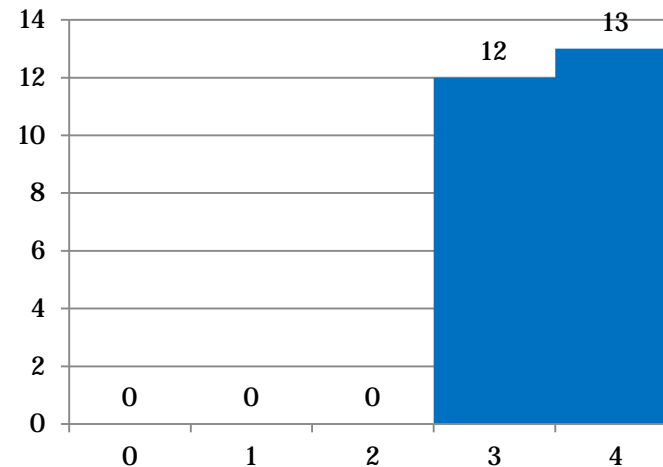


- When the descriptions between score points do not differ very much, scores tend to look like this:

Criteria 1



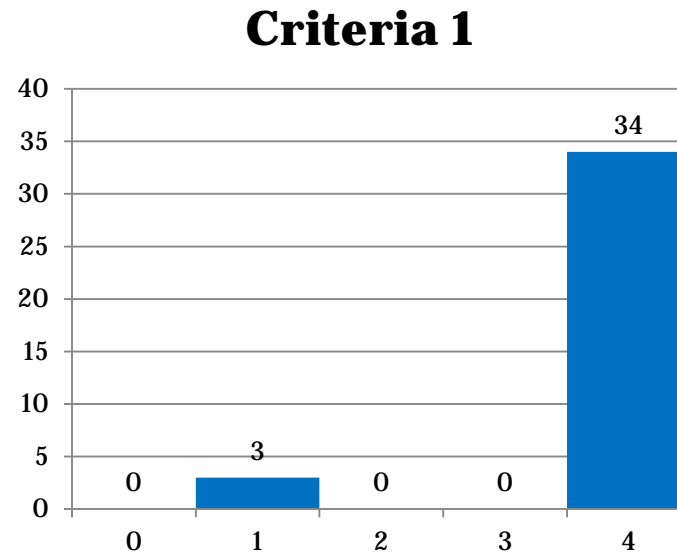
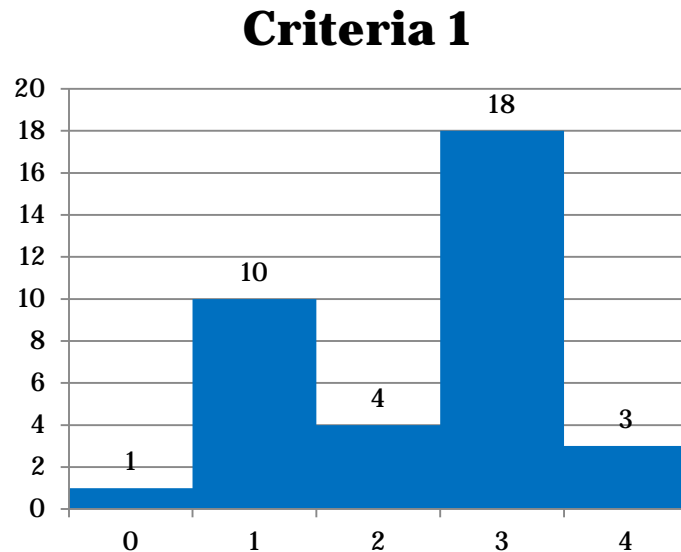
Criteria 2



Criterion-Level Data



- When middle categories on the rubric are not useful, the data looks like this.
 - Reconsider how to rewrite intermediate levels of performance



Criterion-Level Data—Scores of “0”



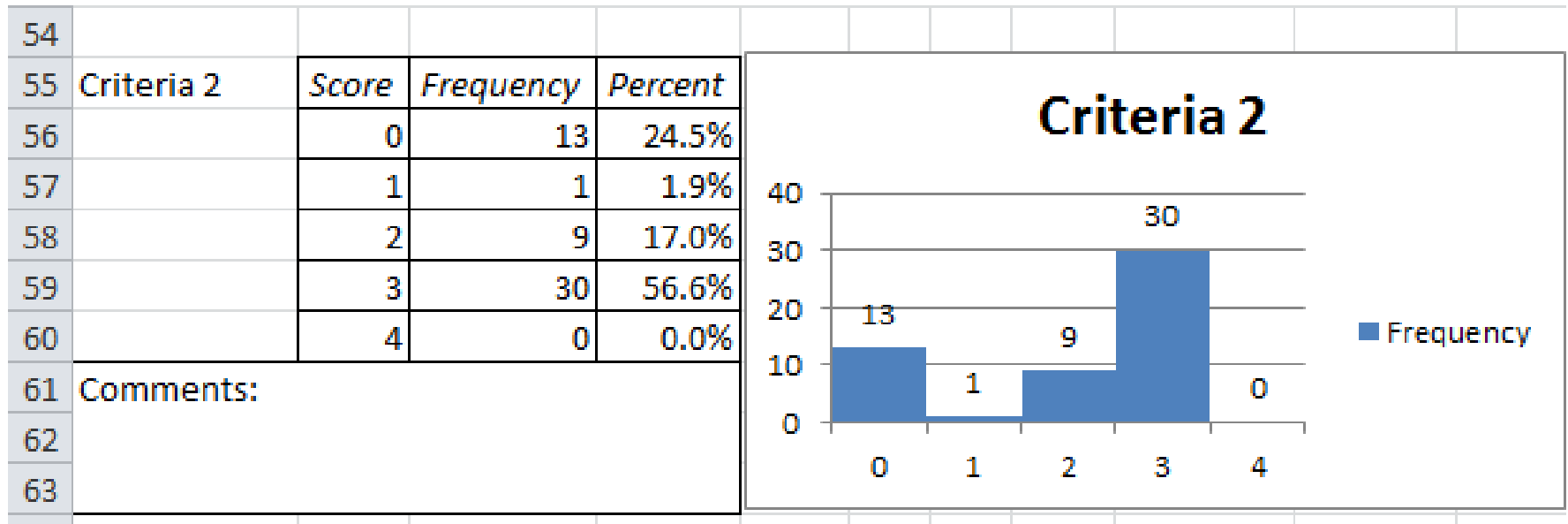
- “Report-main” tab
 - Column I lists the frequency of “0” scores

	A	B	C	D	E	F	G	H	I
1	<u>Content Area</u>		Health						
2	<u>Grade Level</u>		8th Grade						
3	<u>Task</u>		Media Critique and Analysis						
4	Criteria	Description	N	Mean	SD	Min	Max	"0" Scores	
5	1		56	3.07	0.85	0	4	2	
6	2		56	2.00	1.25	0	3	13	
7	3		56	2.30	1.25	0	3	8	
8	4		0						
9	5		0	#####					
10	6		0						
11	7		0						
12	8		0						
13	9		0						
14	10		0						

Criterion Level Data—Scores of “0”



- If there many “0” scores and few/no “1” scores...
 - ...you might need to reevaluate the rubric scale or the task.
 - ...think about why so many students scored “0” on this criterion (see next slide).



Criterion Level Data—Scores of “0”

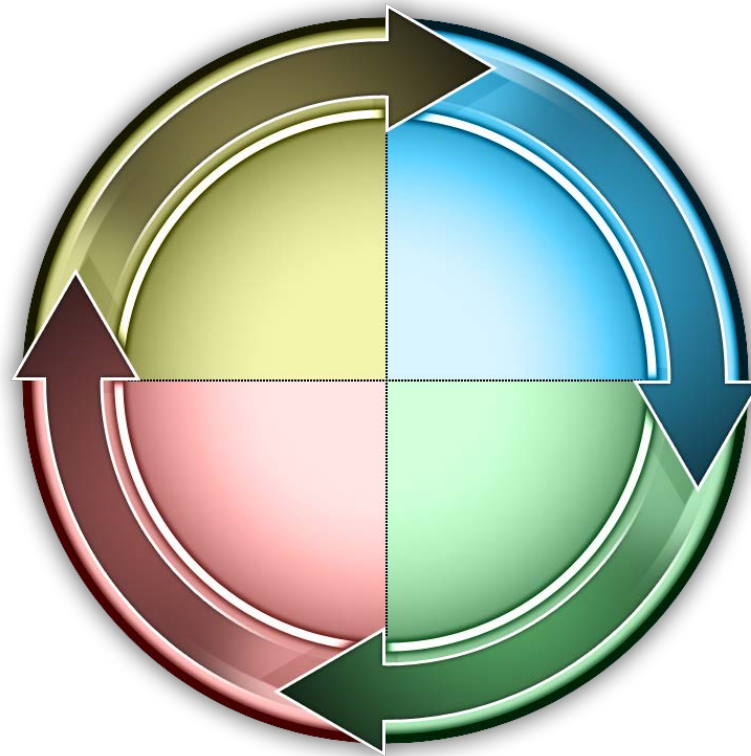


- **Some things to consider:**
 - What does a “0” score indicate about the student?
 - When a “0” score is given, create different codes for this. For example:
 - ✦ Student did not attempt task
 - ✦ Student attempted task but student’s response is not legible
 - ✦ Student’s response is off-topic
 - ✦ Student engaged in unsafe behaviors (CTE)

Revising Tasks & Rubrics



- Revisions of Tasks/Rubrics will not be a sequential process—it will be an iterative one.



Revising Tasks & Rubrics



- **Issue**
 - How can criterion weights (on the rubric) be established?
- **Suggestions:**
 - Weights should be determined by content experts (you!).
 - Weight with percentages. For example, if you have three criteria and the first criterion is most important:

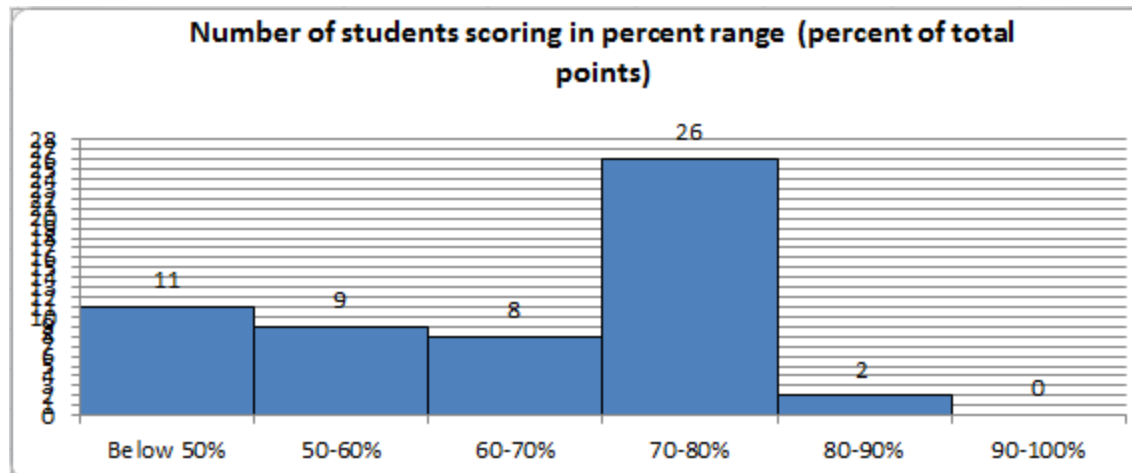
Criterion	Score	Weight			Score
1	4	X	.50	=	2
2	4	X	.25	=	1
3	3	X	.25	=	0.75
				TOTAL	3.75 / 4

Criterion Weights



- You can change the weights to see how this affects your score distributions.
- All criteria weighted equally (3 criteria each weighted 0.33)

Quartile	Range	Frequency	Percent			
1st	0-< 1.98	11	19.6%			
2nd	1.98-< 2.81	17	30.4%	Total (N)	56	
3rd	2.81-< 2.97	0	0.0%			
4th	2.97- 3.3	28	50.0%			



Criterion Weights

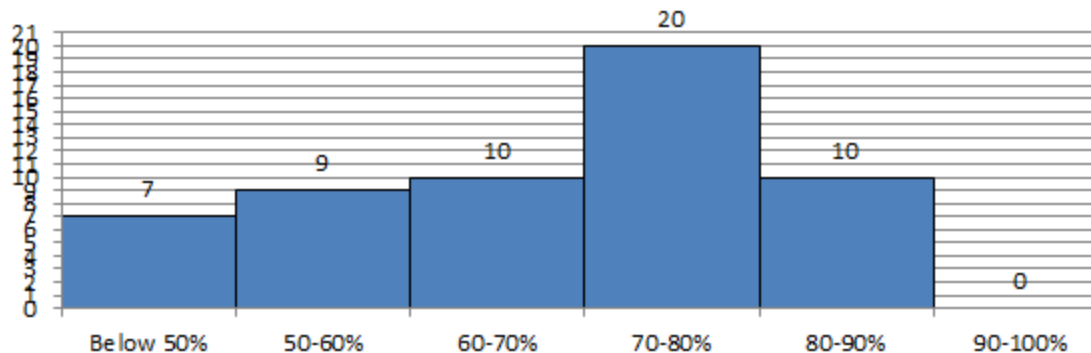


One strong criterion, two weaker criteria:

- Criterion 1 is weighted 0.50
- Criterion 2 is weighted 0.25
- Criterion 3 is weighted 0.25

Quartile	Range	Frequency	Percent			
1st	0-< 2.25	11	19.6%			
2nd	2.25-< 3	15	26.8%	Total (N)	56	
3rd	3-< 3	0	0.0%			
4th	3- 3.5	30	53.6%			

Number of students scoring in percent range (percent of total points)



Criterion Weights

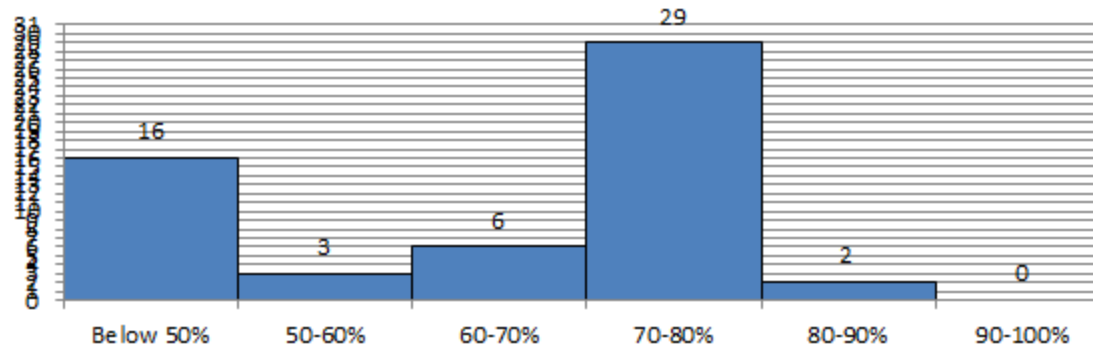


Two strong criteria, one weaker criterion:

- Criterion 1 is 0.20
- Criterion 2 is 0.40
- Criterion 3 is 0.40.

Quartile	Range	Frequency	Percent			
1st	0-< 1.8	12	21.4%			
2nd	1.8-< 2.8	13	23.2%	Total (N)		56
3rd	2.8-< 3	11	19.6%			
4th	3- 3.2	20	35.7%			

Number of students scoring in percent range (percent of total points)



Revising Tasks & Rubrics



- **Issue:**
 - The performance tasks were difficult to administer in the classroom.

- **Suggestions:**
 - Consider feasibility when revising your tasks.
 - ✦ Can the task be simplified for the students/teacher?
 - ✦ If scoring is cumbersome in the classroom setting, should the task be video/audio recorded and scored at a later time?

Revising Tasks & Rubrics



- **Issue:**
 - The skills/knowledge assessed in the tasks do not match the claims you wish to make about the student.
- **Suggestion:**
 - Revise the task so that it elicits evidence of the claims you are trying to make.

Evidence-Centered Design

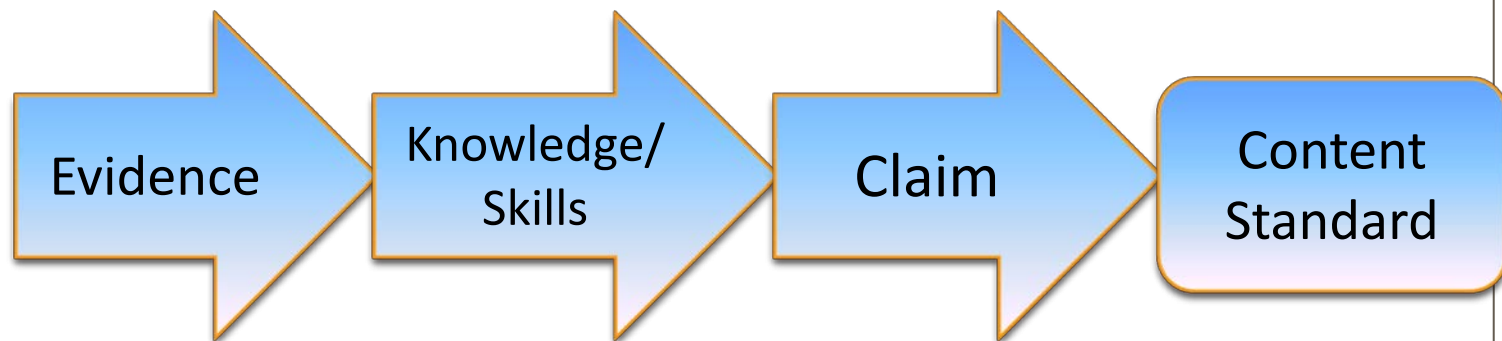


Image Source: Smarter Balanced Assessment Consortium (2012)

Evidence-Centered Design

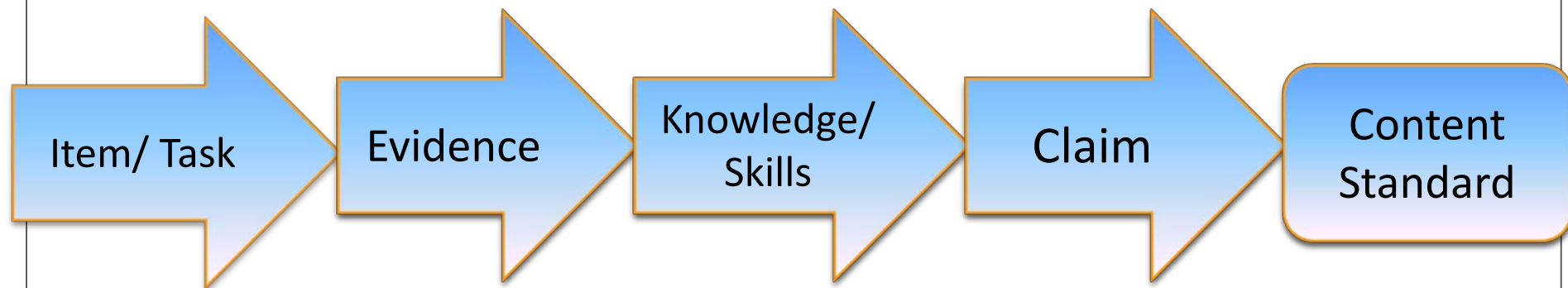
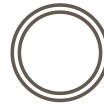


Image Source: Smarter Balanced Assessment Consortium (2012)

Evidence-Centered Design

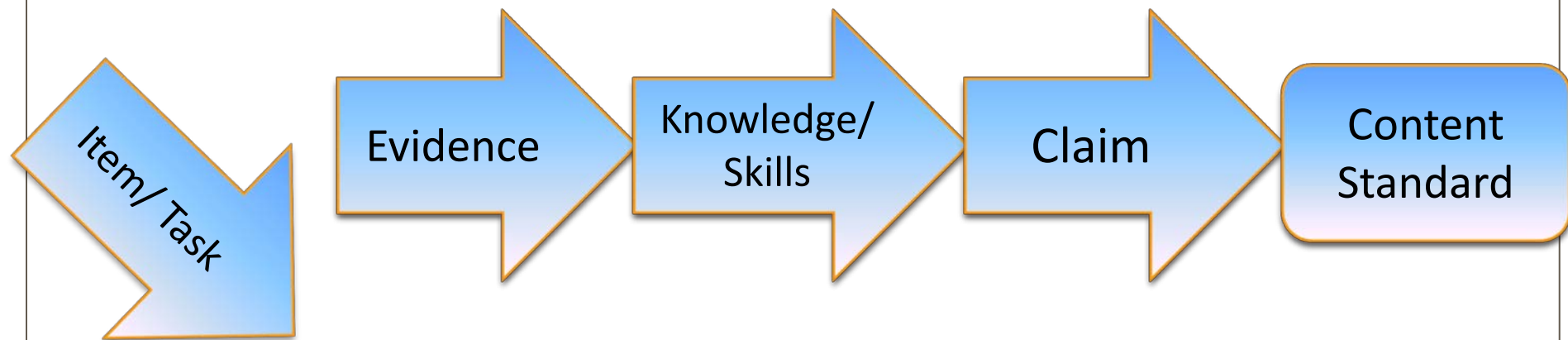


Image Source: Smarter Balanced Assessment Consortium (2012)

Evidence-Centered Design

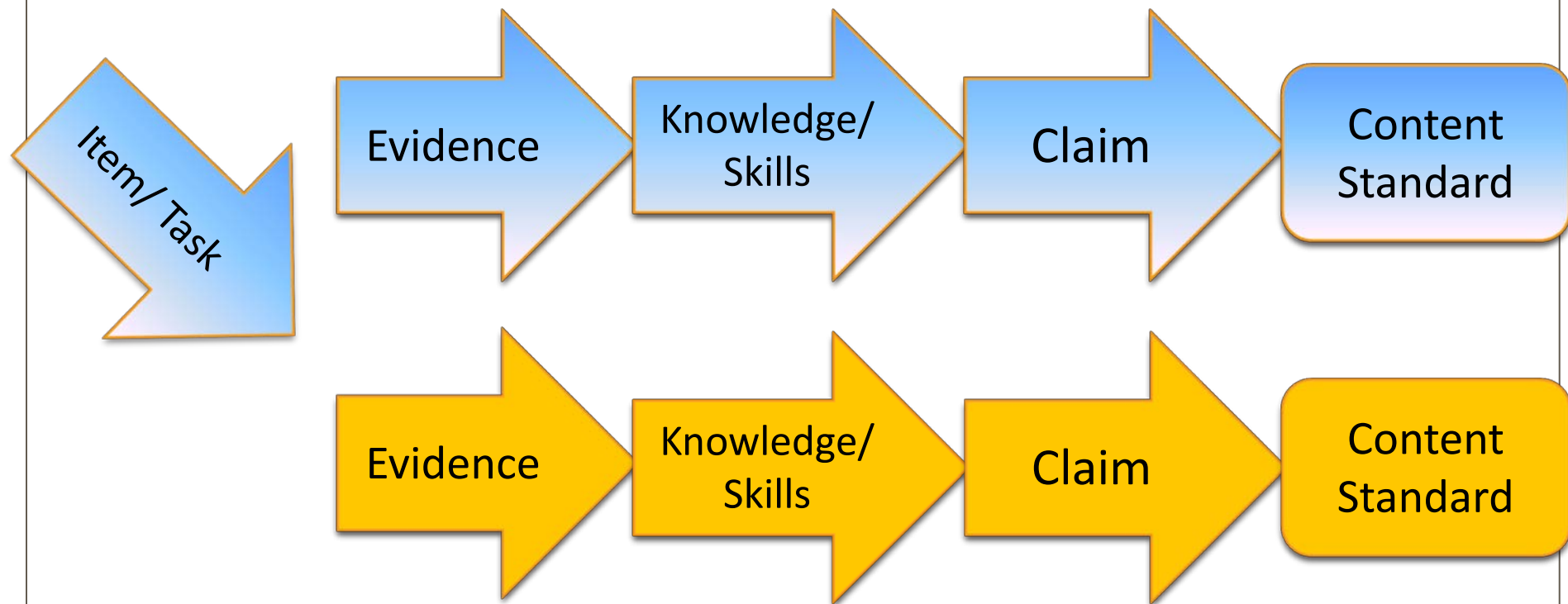


Image Source: Smarter Balanced Assessment Consortium (2012)

Revising Tasks & Rubrics



- **Issue:**
 - Your group decides that the task difficulty level or DOK level is not appropriate given the claims you want to make about the students.

- **Suggestions:**
 - Revise the task to the appropriate difficulty level or DOK level.
 - Revise the rubric accordingly (e.g., simplify the language in the rubric).

Revising Tasks & Rubrics



- **Issue:**
 - Students were confused about the instructions/task.
(and/or)
 - The language in the rubric is too difficult for students to understand.
- **Ask yourselves:**
 - Is the task itself simply too difficult?
 - Are the instructions written at the appropriate grade level?
 - ✦ If not, how can the language be simplified?
 - ✦ Should key concepts be clarified for students?
 - Can you break the task into smaller components?

Revising Tasks & Rubrics



- **Issue:**

- A relevant behavior observed in the student's work/performance did not relate to any criteria in the rubric.

(or)

- A critical behavior described in the rubric was not observed in students' work/performance.

- **Suggestions:**

- Think about whether the behavior relates to knowledge/skills you wish to assess and the claims you wish to make about students.
 - ✦ Do you need to create criteria in the rubric?
 - ✦ Do you need to remove this element from the task or from the rubric?
- Clarify the instructions in the task.

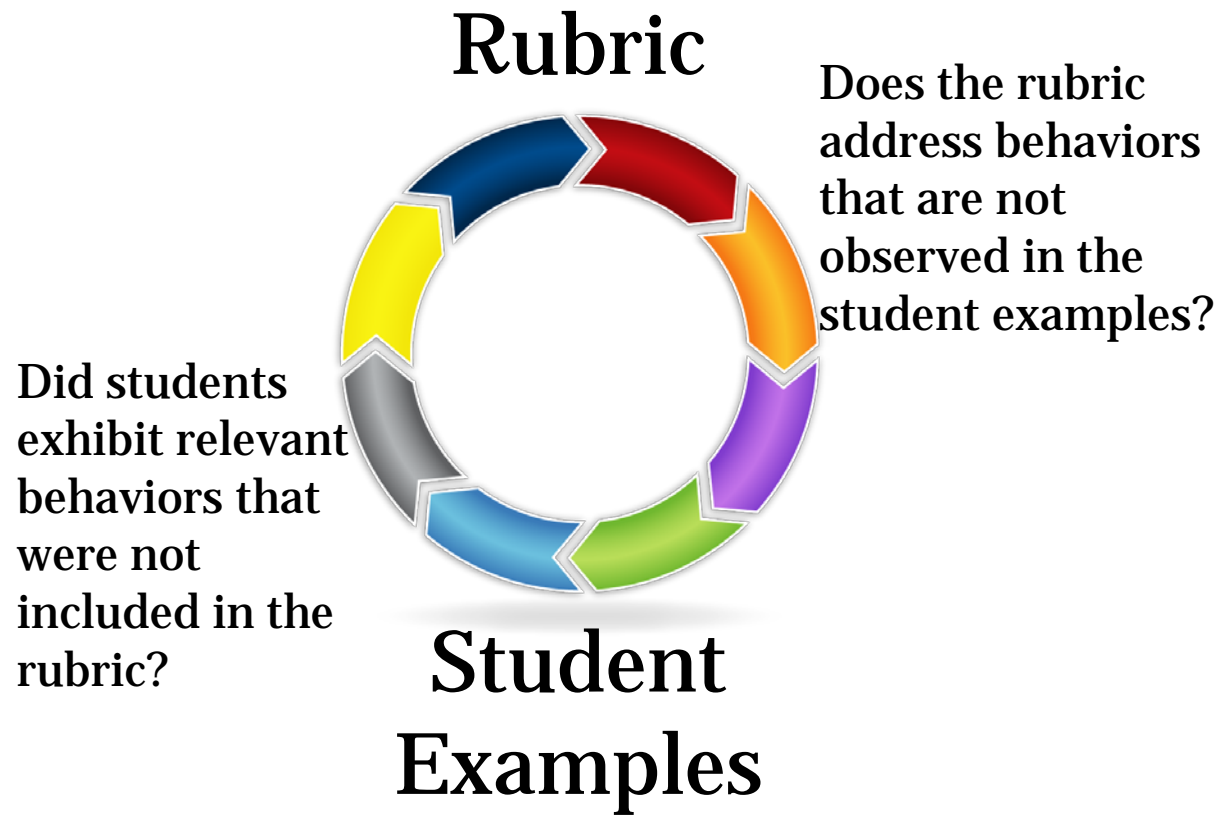
Revising Tasks & Rubrics



- **Issue:**
 - Students were not able to complete task within time constraints.
(OR)
 - Students completed task with a lot of extra time left.

- **Suggestions:**
 - Reevaluate how much time is needed for the task and revise time limits accordingly.
 - If it is not possible to give students more time for the task, and if the task is shortened, make sure you revise:
 - ✦ the rubric
 - ✦ knowledge/skills
 - ✦ claims you are making about students

Rubric Review



Rubric Review—Good Rubrics!



- The criterion is clearly and concisely defined at each score point.

	4	3	2	1
Savings Plan	Report includes thorough summary of savings plan.	Report includes summary of savings plan.	Report includes pieces of a savings plan.	Savings plan is unclear or not included.

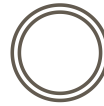
Rubric Review—Good rubrics!



- Behavior is quantified and objective

Scoring Criteria	4	3	2	1
Melodic Improvisation	Melody is in the same tonal center as the call and student responds with 5 or more notes.	Melody has 2-4 notes and stays in the same tonal center as the call.	Melody has 2-4 notes and 2-3 notes are outside of the tonal center.	Melody is on one pitch only or all notes are outside of tonal center.

Rubric Review—Good rubrics!



- Innovative way of identifying Advanced Placement standards of achievement

Scoring Criteria	4	3	2	1
Peer Critique	Provides constructive feedback relevant to this rubric, class goals and objectives.	Analysis and synthesis of peers' performance to complete rubric and justification of their score.	Feedback is short and not concise, lacking support and justifications.	No feedback provided.
Peer Critique Extended	Student provides specific feedback and strategies connected to examples within a performance in a professional and supportive manner.			

Rubric Review—Good rubrics!



- Goals are specific and progressive

Scoring Criteria	4	3	2	1
Shares Poster with Class (Oral Presentation)	Applies content-specific and academic vocabulary to clarify information provided to needs of audience	Uses content-specific and academic vocabulary to provide information to audience	Uses content-specific vocabulary to provide information to audience	Uses basic vocabulary to provide information to audience

Rubric Review



- Issue: Ambiguous language in the rubric point descriptions.

Criterion	4	3	2	1
Identifies Problem & Solution	Both problem and solution are realistic and clear	Problem and solution <u>may be</u> realistic but not clear	Either problem or solution is missing or unrealistic	Both problem and solution are either missing or unrealistic

Rubric Review



- Issue: Ambiguous language in the rubric point descriptions.

Criterion	4	3	2	1
Identifies Problem & Solution	Both problem and solution are realistic and clear	Both problem and solution are may be realistic but one is not clear	Either problem or solution is missing or unrealistic or missing	Both problem and solution are either missing or unrealistic or missing

Rubric Review



- Issue: Rubric point descriptions are not clearly defined.

Criterion	4	3	2	1	0
Connection of drill to skill development plan	<u>Thoroughly</u> explains connection of drill to skill development plan	<u>Mostly</u> explains connection of drill to skill development plan	<u>Minimally</u> explains connection of drill to skill development plan	Has not explained the connection of drill to the skill development plan	Connection to the development plan is not present

Rubric Review



- Issue: Rubric point descriptions are not clearly defined.

Criterion	4	3	2	1
Written report	The written report includes correct use of <u>ample</u> content vocabulary accompanied with correct spelling and grammar.	The written report includes the correct use of <u>some</u> content vocabulary necessary for the task.	The written report includes content vocabulary but uses it incorrectly.	The written report does not include content vocabulary.

Rubric Review



- **Issue: Rubric point descriptions are subjective.**
 - In this case, the teacher is asked to assess his/her response to the student's performance rather than the student's behavior.

Criterion	4	3	2	1
Aesthetic Appeal	Response has a strong aesthetic appeal and general impression	Response includes interesting ideas	Response does hold the listener's interest	Response does not hold the listener's interest

Rubric Review



- Issue: Too much information clustered under one criterion.

Scoring Criteria	4	3	2	1
Conducts Research	Analyze information from a variety of multimedia sources to answer all provided questions with elaborated details, identifies multiple questions of interest to answer	Locates and uses information from a variety of multimedia sources to answer all provided questions with key details, Identifies a specific question of interest to answer	Locates and uses information from one multimedia source to answer some of the provided questions with minimal details, identifies a question to answer	Does not locate or use information from multimedia sources, does not answer any of the provided questions and/or includes irrelevant details, does not identify a relevant question to answer

Rubric Review



Scoring Criteria	4	Scoring Criteria
Conducts Research	Analyze information from a variety of multimedia sources to answer all provided questions with elaborated details, identifies multiple questions of interest to answer	1 Conducts Research
		2 Responds to Questions and Generates Questions

Rubric Review



- Issue: Positive language may hinder the distinction between categories.
 - What is the difference between these two categories?

Criterion	4	3
Creates an organizational structure	<p>Uses logical principles to organize reasons and evidence.</p> <p>The writer has considered the order of the reasons and evidence and how they will be grouped.</p>	<p>Organizes reasons and evidence clearly. It is clear how evidence connects and corresponds to the main idea.</p> <p>Uses paragraphing and makes other structural decisions (choice of linking words, order of information, etc.) to organize reasons and supporting evidence.</p>