

Construction and Evaluation of Test Items: Part 2

Yudariah Mohammad Yusof
Mitra Mohd Addi
Centre for Engineering Education, UTM

UTeM @20April2021



Session 2

Rubrics in Assessment

Intended Learning Outcomes

At the end of the session participants should be able to

- Evaluate test items and examination questions
- Analyze and plan the implementation of scoring

Rubrics “are designed to function as scoring guidelines, but they also serve as arbiters of quality and agents of control” over what is taught and valued.

Mabry, 1999

Types of Scoring Instruments



Exam: Marking Approaches

- Using assessment criteria
 - describe the knowledge and skills that a student is expected to demonstrate
 - set the standards - describe the different bandings of overall achievement
 - provide an objective basis for interpreting and differentiating student performance
 - more generic and used as a framework to fairly judge the merits of each student's work across a whole course
- Using marking schemes
 - a set of specific marking guidance
 - vary with each and every question and are tailored and specific

Marking Guidance

- Writing the exam question is only half the story – producing the associated marking guidance must also be considered.
- Marking guidance can take a number of different forms ranging from specimen or model answers through to descriptive criteria and detailed marking schemes matched to necessary answer content.

Example

Source: Yahya, 2013

(b) *Mavi Marmara II* (the same ship as in Question 1), LBP 126m, Breadth 24m, Depth 11.0 m and displacement 15000 tonnes, KB and KG is 3.9 m and 7.35 m above keel respectively. During one of her humanitarian voyage, while the ship is floating at draft 6.45m, one of the forward compartment at port was damaged by the external forces. Based on the GA drawing the damaged compartment has the following particulars; Average waterplane size, length 11m x width 10m, centroid at 27m forward of amidships, 2.5m from Centre line on port side. The compartment can be approximately assumed having a cuboid shape with the centre of volume at 0.5 draft and permeability for volume and area 0.65. Determine whether the ship is still survive or not under this damaged condition.

15 marks

Marking Scheme

| STEP | ACTIVITY | MARK |
|------|---|-----------|
| 1 | Calculate Parallel Sinkage | 1 |
| 2 | Calculate Trim | 5 |
| 3 | Calculate Draft at CL due to Trim Only | 1 |
| 4 | Calculate Draft at Side due to heel | 5 |
| 5 | Calculate new GM (min Stability) | 2 |
| 6 | Overall Judgement on ship survivability | 1 |
| | Total Mark | 15 |

Question (b)

LC = 11 Bc = 10 Tc = 6.45
 LCF = 27 TCF = -2.5 To Port Rho =

STEP 1 : Cal Parallel Singkage

Vc = 461.175
 Aw = 2128.00
 ua = 71.5
 Parallel Singkage = 0.22 m

Marking Scheme

STEP 2 : Trim

(a) - Calculate P

| Comp | Area | LCF | ML |
|--------------|---------|-------|---------|
| Full Area | 2128.00 | -3.32 | -7056 |
| Damaged Comp | -71.5 | 27.00 | -1930.5 |
| | 2056.50 | | -8986.5 |

F1 = -4.37 m from Midship

P = 31.37 m

(b) - Cal Moment Trim

Vc = 461.175 m³

Moment Trim = 14828.6 Tonnes-m

(c) - Calculate ILF1 after damaged

ILCF before damaged = 1.75E+06 m⁴

ILF1 before damaged = 1.76E+06

ILF of Comp at its own Centroid = 7.21E+02 m⁴

IF1 of Damaged Comp = 7.11E+04 m⁴

ILF1 after damaged = 1.69E+06 m⁴

(d) - Calculate MCTcm

ILF1 = 1.69E+06 m⁴

MCTcm = 137.15 Tonnes-m

(e) - Cal Trim

Trim = 1.08 m

STEP 5 : GM changes

(a) - GM before Damaged

KBo = 3.90 m
 Vol Full = 14634.15 m³
 BMo = 5.09 m
 KGo = 7.35
 GMo = 1.64 m

(b) - GM After Damaged

KBn = 4.01 m
 ITn = 73375.09 m⁴
 BMn = 5.01 m
 KGn = 7.35
 GMn = 1.67

| Comp | Vol | KB | Mv |
|-------------|----------|-------|----------|
| Ship Intact | 14634.15 | 3.9 | 57073.17 |
| Comp | -461.18 | 3.225 | -1487.29 |
| Layer add | 461.18 | 6.56 | 3026.29 |
| after Dam | 14634.15 | 4.01 | 58612.17 |

Changes in GM = 0.03 m Better Stability

STE

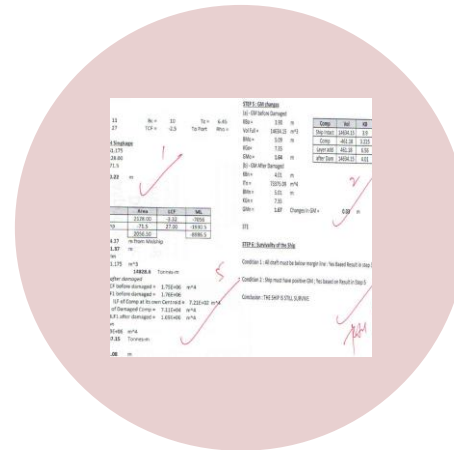
STEP 6 : Survivability of the Ship

Condition 1 : All draft must be below margin line : Yes Based Result in step 3 & 4

Condition 2 : Ship must have positive GM ; Yes based on Result in Step 5

Conclusion : THE SHIP IS STILL SURVIVE

Activity



Scrutinize the given marking scheme.

Identify how the marks are distributed. What do students need to demonstrate in order to get marks?

Example – Marking scheme

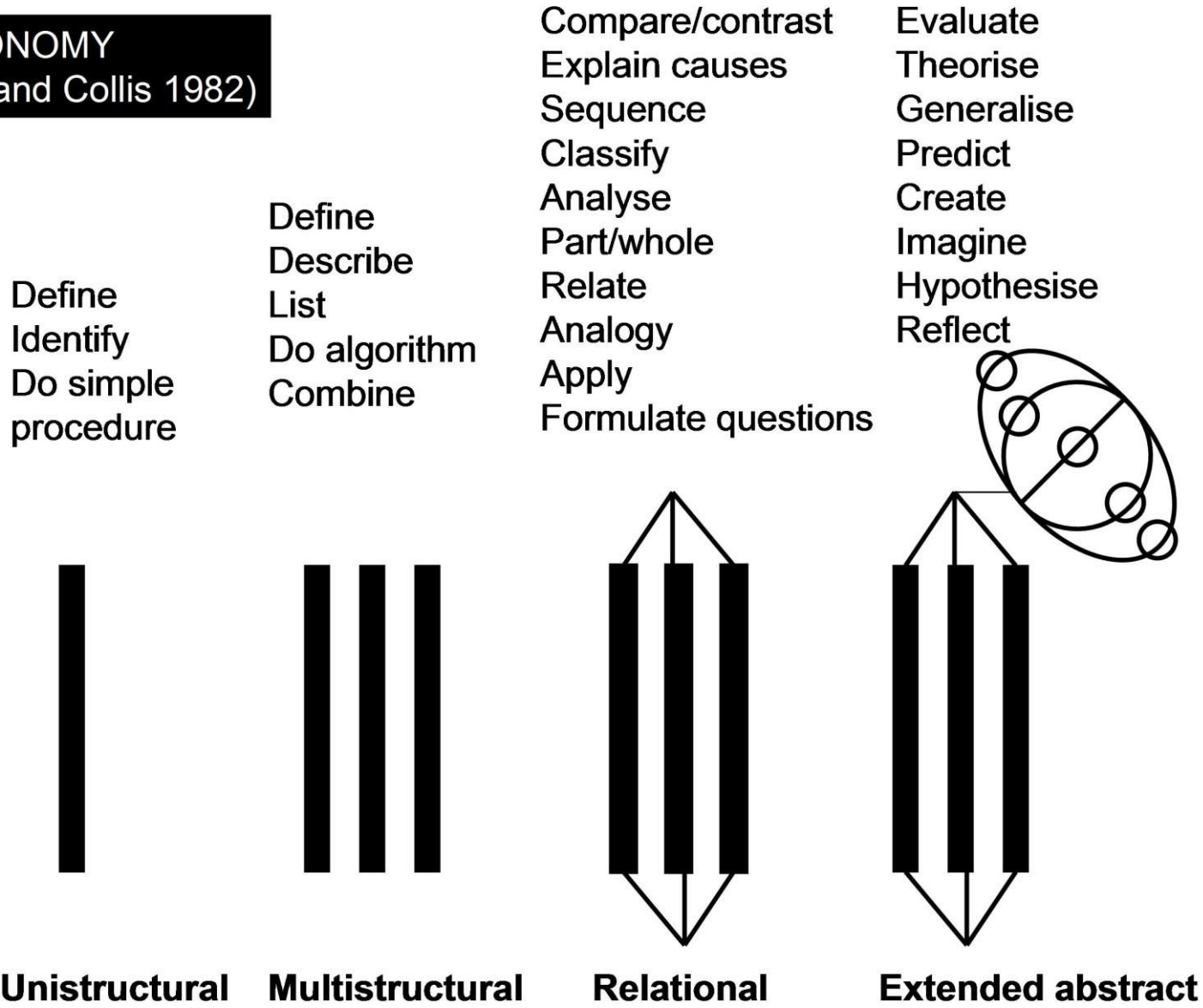
Marks are of following three types:

| Notation | Descriptor | Notes |
|----------|---------------|---|
| M | Method mark | Awarded for a valid method applied to the problem. Method marks are not lost for numerical error, algebraic slips or errors in units. Correct application of a formula without the formula being quoted will earn the M mark. M mark can also be implied from a correct answer. |
| A | Accuracy mark | Awarded for a correct answer or intermediate step correctly obtained. Accuracy marks cannot be given unless the associated method mark is earned or implied |
| B | Result mark | Mark for a correct result or statement independent of method marks. |

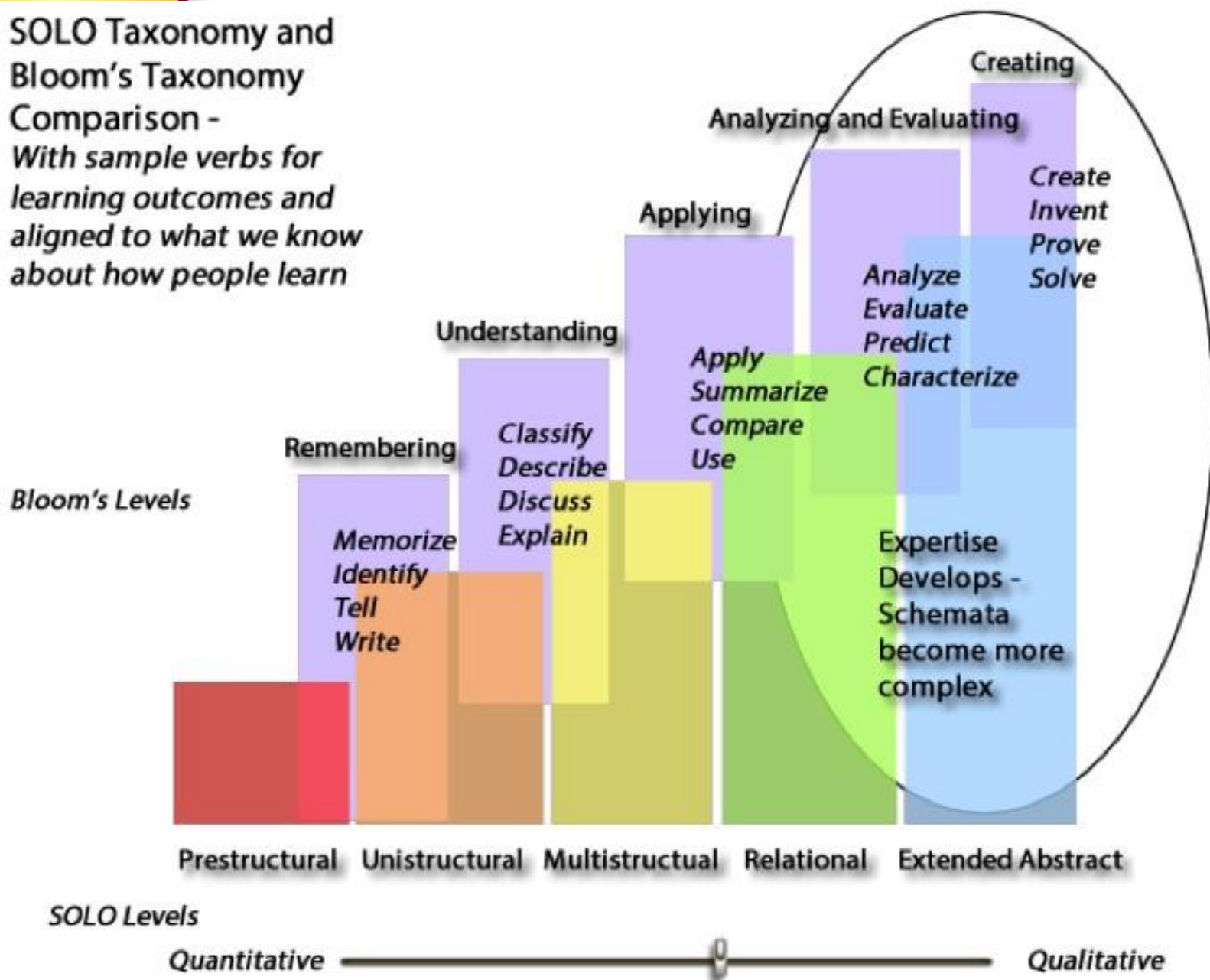
Structure of Observed Learning Outcomes

SOLO TAXONOMY (after Biggs and Collis 1982)

➤ used to
analyse &
categorize
the quality
of students'
responses



SOLO Taxonomy and Bloom's Taxonomy Comparison -
With sample verbs for learning outcomes and aligned to what we know about how people learn



How do We Know LO been Achieved?

Performance – indicators of learning outcomes

- things students would do, make, say, or write
- rubric *describes* the performance.
- Give a structure to provide more effective feedback to students
- Provide valuable information about which aspects of the course are working well and which are not

General Rubrics

- Use criteria and descriptions of performance that *generalize* across or can be used with, different tasks. The tasks all have to be instances of the same learning outcome.
- The criteria point to aspects of the learning outcome and not to features of any one specific task.
- Descriptions focus students on what their learning target/outcome (knowledge & skills) is supposed to be (general qualities).
- Assessing product or process?

Development of Rubrics

Rubrics

- Marking tools – developed hand-in-hand with ATs
- Reflect attention & purpose of assessment
 - must be aligned to your course outcomes and your assessment outcomes
- Bring clarity & transparency to assessment
- Ascertain whether or not students have achieved the learning outcomes – quantify achievement of outcomes
- Communicate to students (and to other assessors) your expectations in the assessment, and what you consider important
- Can be use as teaching resources
 - give students a clear indication of any critical gaps or weaknesses in their performance, what they need to do to improve, and where to focus their efforts (feedforward rather than feedback)

Linking Outcomes & Performance Attributes/Success Criteria

- Criteria focused around knowledge & skills to be assessed.
- verbs in LO describe student action
 - assessment measure action

Example

| LO | I'm looking for... (Verb in LO) | My expectation... (Descriptor of the verb) | To be successful students need to... (Performance attributes/success criteria) |
|----------------|------------------------------------|---|---|
| Explain ... | Explain | Explain concept or scenario etc. | Evaluate concept or scenario with clarity providing reasoning for your statements with examples |



| Criteria | 4 | 3 | 2 | 1 |
|----------------------------------|---|---|--|---|
| Explain concept or scenario etc. | Evaluate concept or scenario with clarity providing reasoning for your statements with examples | Analyze concept or scenario etc. with clarity providing reasoning for your statements | Explain concept or scenario etc. with clarity providing reasoning for statements | Explain concept or scenario etc. with clarity |

(b) *Mavi Marmara II* (the same ship as in Question 1), LBP 126m, Breadth 24m, Depth 11.0 m and displacement 15000 tonnes, KB and KG is 3.9 m and 7.35 m above keel respectively. During one of her humanitarian voyage, while the ship is floating at draft 6.45m, one of the forward compartment at port was damaged by the external forces. Based on the GA drawing the damaged compartment has the following particulars; Average waterplane size, length 11m x width 10m, centroid at 27m forward of amidships, 2.5m from Centre line on port side. The compartment can be approximately assumed having a cuboid shape with the centre of volume at 0.5 draft and permeability for volume and area 0.65. Determine whether the ship is still survive or not under this damaged condition.

15 marks

Answering Process (Expectation)

1. Reading and understanding the question
2. **Identifying the problem** to be solved
(Parameters to be calculated)
3. Highlighting all the **information / data given** – relate them to the problem posed
4. Planning the **calculation procedure, step by step**
5. For each step, recalling the **appropriate formula / theory / concept** to be used
6. **Performing the calculation**, using the appropriate formula and making full used of the data given
7. **Checking the accuracy** of the calculation
8. **Writing the final answer** in appropriate format
(one unique answer)
9. Writing the **conclusion** (Survive or not) based on the result

Using Rubrics

| NO | ABILITY | MARK |
|-------------------|---|-----------|
| 1 | Identify the problem correctly | 2 |
| 2 | The problem solved using correct procedure, step by step | 5 |
| 3 | For each calculation step, appropriate formula / concept applied using appropriate data given | 3 |
| 4 | Calculation performed accurately | 2 |
| 5 | Results presented according to appropriate format | 1 |
| 6 | Technical judgment correctly made based on the result | 2 |
| TOTAL MARK | | 15 |



| ASSESSMENT CRITERIA | SCORE |
|---|-------|
| All 6 steps carried out with correct sequence | 5 |
| Some step missing / not shown or not in proper sequence | 3-4 |
| Most step missing | 1-2 |
| No calculation step shown | 0 |



| ASSESSMENT CRITERIA | SCORE |
|---|-------|
| Judgment made correctly | 2 |
| Judgment made but not align to the result | 1 |
| No judgment or comment made | 0 |

Marking scheme with rubrics – tend to give better result and provide essential information on students' true ability (strength or weaknesses) that lead to appropriate CQI effort

Types of rubrics

Four commonly used types:

- ▶ Checklists
- ▶ Rating scales
- ▶ Holistic rubrics
- ▶ Descriptive/Analytic rubrics

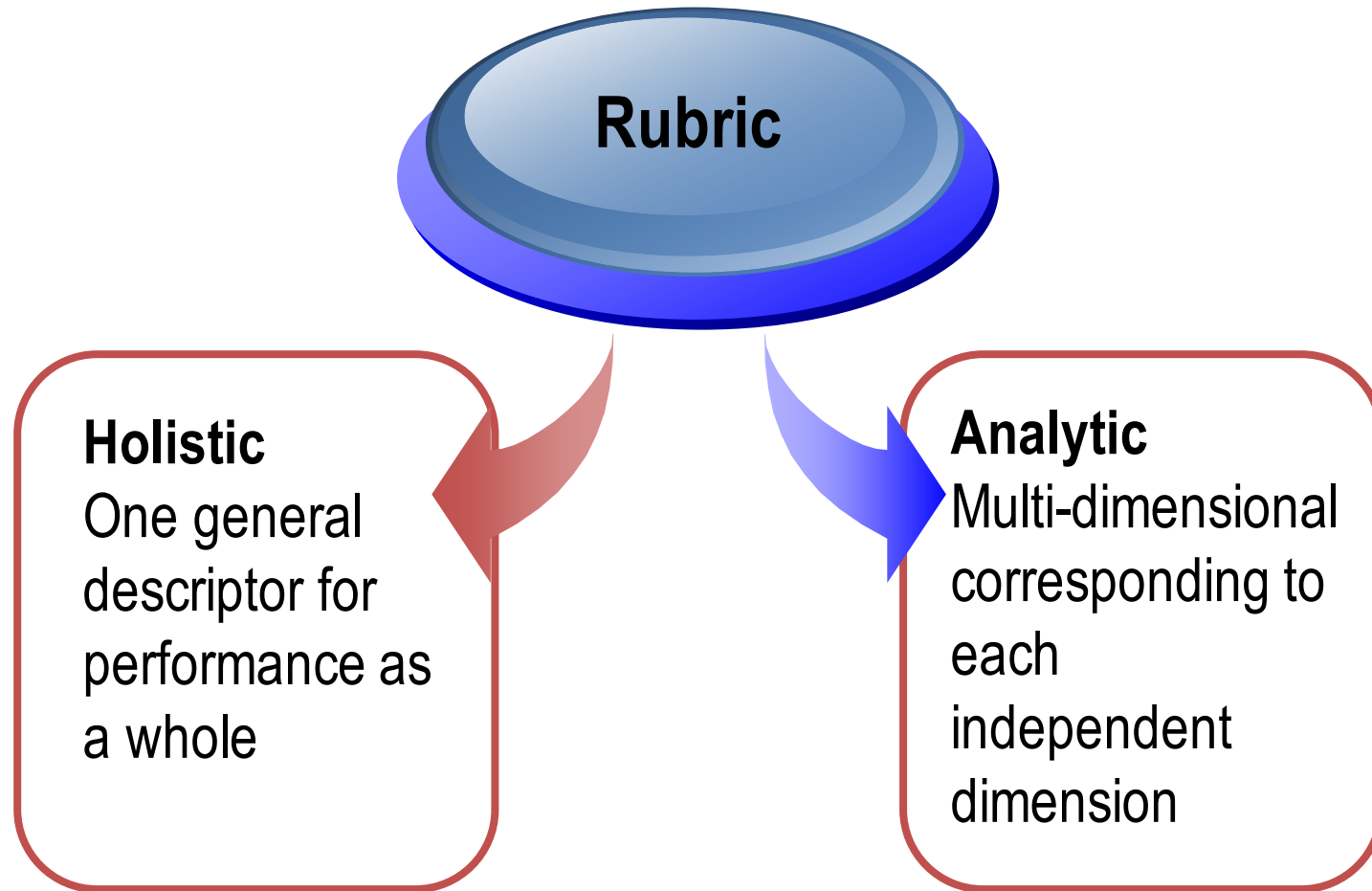


Rating Scales – Example Problem Solving

Rating Form

| | Criteria | Score |
|----|---------------------------------------|--------------|
| 1. | Identify problem, question, or issue | |
| 2. | Consider context and assumptions | |
| 3. | Develop own position or hypothesis | |
| 4. | Present and analyse supporting data | |
| 5. | Integrate other perspectives | |
| 6. | Identify conclusions and implications | |
| 7. | Communicate effectively | |

Holistic vs Analytic Rubrics



Parts of a Rubric

Criteria

**Knowledge/
skills to be
assessed**

Indicators

**Evidence for
judging
degree of
criteria
mastery**

Scoring Scales

**Discriminate
levels of
performance**

Example

| LO | I'm looking for... (Verb in LO) | My expectation... (Descriptor of the verb) | To be successful students need to... (Performance attributes/success criteria) |
|--|------------------------------------|--|--|
| Discuss the topic, context or situation... | Discuss | Discuss all aspects of the topic (facts, principles, rules theories) in group/pair discussion and in class discussion actively | <ul style="list-style-type: none"> Participates actively in group/pair discussion. Participates actively in class discussion <ul style="list-style-type: none"> - Voices views actively without prompting. - Responds actively to viewpoints raised by other students in class. - Elaborates well on viewpoints. |



| Criteria | 4 | 3 | 2 | 1 |
|--|---|---|---|---|
| <ul style="list-style-type: none"> Participate in group/pair discussion Participate in class discussion. | <p>Participates actively in group or pair discussion.</p> <p>Voices views actively without prompting.</p> <p>Responds actively to viewpoints raised by other students in class.</p> <p>Elaborates well on viewpoints.</p> | | | <p>Hardly participates in group/pair discussion; does not show interest in ongoing discussion.</p> <p>Hardly voices views even when prompted.</p> <p>Hardly responds to viewpoints raised by other students in class even when prompted; does not take an interest in ongoing class discussion.</p> |

Example – Holistic rubrics

Assessment Guidelines for Classroom Discussion (10%)

Source: Khairi Izwan, 2016

| Grade | Descriptors |
|-------------------------|---|
| A- to A+ (7.5 – 10) | <ul style="list-style-type: none"> ● Participation in group/pair discussion <ul style="list-style-type: none"> - Participates actively in group/pair discussion. ● Participation in class discussion. <ul style="list-style-type: none"> - Voices views actively without prompting. - Responds actively to viewpoints raised by other students in class. - Elaborates well on viewpoints. |
| B- to B+ (6.0 – 7.4) | <ul style="list-style-type: none"> ● Participation in group/pair discussion <ul style="list-style-type: none"> - Participates in group/pair discussion. ● Participation in class discussion <ul style="list-style-type: none"> - Voices views with occasional prompting. - Responds to viewpoints raised by other students in class. - Elaborates on viewpoints. |
| C- to C+ (4.5 – 5.9) | <ul style="list-style-type: none"> ● Participation in group/pair discussion <ul style="list-style-type: none"> - Participates in group/pair discussion, but participation is inconsistent, with occasional lack of interest. ● Participation in class discussion <ul style="list-style-type: none"> - Voices views but requires frequent prompting. - Responds to viewpoints raised by other students in class but requires frequent prompting. - Occasionally elaborates on viewpoints. |
| D+ (4.0 – 4.4) | <ul style="list-style-type: none"> ● Participation in group/pair discussion <ul style="list-style-type: none"> - Hardly participates in group/pair discussion; generally lacks interest in ongoing discussion. ● Participation in class discussion <ul style="list-style-type: none"> - Voices views only when prompted. - Responds to viewpoints raised by other students in class only when prompted; generally lacks interest in ongoing class discussion. - Hardly elaborates on viewpoints; contribution is limited mainly to short utterances or phrases. |
| D, D- , E (0 – 3.9) | <ul style="list-style-type: none"> ● Participation in group/pair discussion <ul style="list-style-type: none"> - Hardly participates in group/pair discussion; does not show interest in ongoing discussion. ● Participation in class discussion <ul style="list-style-type: none"> - Hardly voices views even when prompted. - Hardly responds to viewpoints raised by other students in class even when prompted; does not take an interest in ongoing class discussion. |

Activity

Suppose you have a CLO with the active verb discuss/distinguish/differentiate and you aim to give a written test item.

Identify the performance attributes/success criteria.

Analytic Rubric

The following is a simple quantitative approach to assigning marks/points to each criterion. Score all items on a 4-point scale:

4 – Task or element done

3 – Task or element done, but only partially correct

2 – Task or element done, but clearly incorrect

1 – Task or element not done

➤ 1s, 2s and 4s will be clearly identifiable, anything else is a 3.

Example – Analytic rubrics

| Evaluation Criteria | 4 | 3 | 2 | 1 |
|---|--|--|--|--|
| Attends group meetings regularly and arrives on time | Attended 100% of the group meetings | Attended 60% of the group meetings | Attended 30% of the group meetings | Fails to attend the group meetings |
| Contributes useful ideas | Consistently gives more ideas than expected | Occasionally gives ideas to the group | Sometime gives ideas to the group | Gives no ideas |
| Participates in group discussions | Consistently exceeds expectations for participation and contributes relevant material to project | Regularly participate in group discussions and share relevant material | Sometime participate in group discussions but rarely share relevant material | Fails to participate in group discussions and fails to share relevant material |
| Completes assigned task on time | Consistently carries out assigned tasks and always volunteers for other tasks | Regularly carries out assigned tasks and occasionally volunteers for other tasks | Sometime carries out assigned tasks but never volunteers to do a task | Unwilling to carry out assigned tasks |
| Quality manner of completed work | Produces work that exceeds group and project expectations | Meets minimum group and project expectations | Occasionally meets minimum group and project expectations | Produces unacceptable work, fails to meet group and project requirements |
| Contributes significantly to the success of the project. | Contribution exceeds group requirements | Contribution meets group requirements | Contribution fails to meet some group requirements | Fails to meet group requirements |

Evaluating your Rubrics

- **Does the rubric relate to the outcome(s) being measured?**
 - address the criteria of the outcome(s) to be measured
- **Does it cover important criteria for student performance?**
 - authentic, reflect what was emphasized for the learning outcome and assignment(s)
- **Does the top end of the rubric reflect excellence?**
 - acceptable work clearly defined; high point on the scale truly represent excellent ability
- **Are the criteria and scales well-defined?**
 - clear scale for each criterion measures; possible to easily differentiate between scale points; tested with actual student products to ensure all likely criteria included
- **Can the rubric be applied consistently by different scorers?**
 - Inter-rater reliability (also called inter-rater agreement), shared understandings of definitions and criteria, scales, and descriptors increase consistency.

Template for Analytic Rubrics

| | Beginning 1 | Developing 2 | Accomplished 3 | Exemplary 4 | Score |
|--------------------|---|---|--|---|--------------|
| Criteria #1 | Description reflecting beginning level of performance | Description reflecting movement toward mastery level of performance | Description reflecting achievement of mastery level of performance | Description reflecting highest level of performance | |
| Criteria #2 | Description reflecting beginning level of performance | Description reflecting movement toward mastery level of performance | Description reflecting achievement of mastery level of performance | Description reflecting highest level of performance | |
| Criteria #3 | Description reflecting beginning level of performance | Description reflecting movement toward mastery level of performance | Description reflecting achievement of mastery level of performance | Description reflecting highest level of performance | |
| Criteria #4 | Description reflecting beginning level of performance | Description reflecting movement toward mastery level of performance | Description reflecting achievement of mastery level of performance | Description reflecting highest level of performance | |

Assessing Test Item Quality

Several methods for collecting feedback on the quality of your test items.

- Use self-review checklists
- Student evaluation of test item quality
- Vetting by a panel/committee
 - format, cognitive demand, content validity, editorial review (spelling, grammar, etc.), sensitivity, fairness

Checklist for Evaluating Test Items – Example Essay

_____ Prepared items that elicited the type of behaviour I wanted to measure.

_____ Phrased each item so that the student's task was clearly indicated.

_____ Indicated for each item a point value or weight and an estimated time limit for answering.

_____ Asked questions that elicited responses on which experts could agree that one answer is better than others.

_____ Avoided giving the student a choice among optional items.

_____ Administered several short-answer items rather than 1 or 2 extended-response items.

Student Evaluation of Test Item Quality – Example Questionnaire

| Questions | Response | |
|--|-----------------------|---------------------|
| How would you rate the examination questions? | Excellent | Poor |
| How well did examination questions reflect content and emphasis of the course? | Well related | Poorly related |
| The exams reflected important points in the reading assignments. | Strongly agree | Strongly disagree |
| Examinations mainly tested trivia. | Strongly agree | Strongly disagree |
| Did the exams challenge you to do original thinking? | Yes, very challenging | No, not challenging |
| How difficult were the examinations? | Too difficult | Too easy |

Example – Vetting Checklist

| Checklist | <u>Lecturer / Course Coordinator</u> Tick (✓) | <u>Vetter 1</u> Tick (✓) | <u>Vetter 2</u> Tick (✓) |
|---|--|-----------------------------|-----------------------------|
| 1. Check the formatting using standard template. | | | |
| 2. Check whether the questions fulfill the course outcomes according to Bloom's Taxonomy. | | | |
| 3. Balanced levels of assessment in final examination. | | | |
| 4. Check the coverage of the courses. | | | |
| 5. Check the number of questions with respect to the duration. | | | |
| 6. Check the spelling and grammatical errors. | | | |
| 7. Check all the availability of required data (please refer to the marking scheme). | | | |
| 8. Check the figures and diagrams (readability, clearness, accuracy, aspect ratio, sizes etc.). | | | |
| 9. Is there any complex problems addressed in the questions? | | | |

Activity

Consider the test item.

- What do students actually have to do to get rewarded?
- What was most valued in the item?
 - what gave students best return for effort
- What was least valued in the item?
 - what gave students least return for effort
- What were the assumed knowledge and skills required for the task?

A company, SubSea Malaysia Bhd. intend to exhibit its capability in building and maintaining submarine to the public. For this reason, SubSea decided to hire a used submarine from abroad and display it alongside the yard. The Minister of Defense and his advisory committee is scheduled to visit the yard and eventually decide whether SubSea will be awarded the submarine project or not. For some reason, one of the submarine's compartment is leaking and the vessel tilt (trim) and heel to one side. This could give negative impression to the Minister and his team. You have been given the task to solve the problem within less than 24 hours, in which repairing the damaged compartment is not possible. Put forward your suggestion and support it with appropriate technical arguments based on the knowledge that you have learnt in the course. (15 marks)

Source: Yahya, 2016

Activity: Rubrics

1. Select one question from your Final Exam Script
2. Identify (list) the learning behaviors or performance expectations that you hope to see.
3. Prepare the Assessment Rubrics.

What rubrics can and cannot do...

- Rubrics **does not** make the assessment decision; this is the responsibility of the assessor.
- Rubrics provides a guiding frame for focusing attention on the key elements/constructs (performance criteria) of the assessment area and summary descriptors of a range of performances.

Reflection

- One Take-Away
- One Surprise
- One Area to Explore Further...





Thank you