Ministry of Higher Education and Scientific Research

Faculty of Letters and Foreign Languages

University of Abbes Laghrour- Khenchela

Department of English Language and Literature

Time Allotted: 1:30h

Student's name: **Level: Master1Didactics**

The First Term Exam of Assessment Methods

In a well-structured essay, **answer one of the following questions:**

1. Compare **validity** and **reliability** in traditional vs. alternative assessment: which one is usually easier to secure, and why?
2. If two raters disagree on a score, what steps would you take to diagnose whether the problem is the rubric, the raters, or the task?
3. Analyze the difference between assessing the **product** (final presentation/essay) and the **process** (drafting/rehearsal). Which is more meaningful and why?

……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……
……

……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
………………………………………………………………………….. G.Luck.

Ministry of Higher Education and Scientific Research

Faculty of Letters and Foreign Languages

University of Abbes Laghrour- Khenchela

Department of English Language and Literature

The First Term Exam Master Correction of Assessment Methods

1) Traditional vs. alternative assessment: The argument about validity vs. the one about reliability, respectively

Question: In traditionaland alternative assessment, compare validity and reliability;in general which is easier to obtain, Why?

Master correction (model answer)

Validity refers to how well a test measures what it says it will measure (e.g. communicative competence, argumentative writing). Reliability is the degree of repeated administration (for a different time, task or rater).

It is much easier to reliably assess in traditionel testing (multiple choice, short answers, standardized presentations) because:

structured and standardized tasks,

• scoring is frequently objective or rule driven,

• conditions are manipulable (same timing of testing, same items, same scoring key),

•It is easier to have statistical consistency.

Traditional tests, however, can face issues of validity when assessing complex constructs such as speaking interaction or authentic writing in the real world because they frequently excerpt language from context (Chapelle 1998).

Validity is frequently easier to enhance in alternative assessment (e.g., portfolios, projects, presentations, performance tasks) because:

• activities are more authentic and connected to actual language use,

• they may represent intricate constructs (interaction, discourse, pragmatic competence),

• evidence can be more extensive and indicative of course outcomes.

But other forms of testing many times fail reliability tests with:

• tasks are open-ended,

• scoring involves human judgment,

• performance depends on the topic, the group dynamics or even the context,

• results may be affected by rater differences (severity/leniency).

Conclusion: Traditional assessment is more consistently reliable in general and alternative assessment supports validity better, for the most part—unless strong rubrics, rater training/marker standardization/moderation practices and task standardization are put into place to drive up the reliability of alternative formats.

2) Two raters dissagree: rubric, raters or task?

Question: If you've got two raters disagreeing about a score for an observation, what do you to diagnose the root cause – is it the rubric? Is it the rater? Or is it task?

Master correction (model answer)

For each one I wanted to diagnose disagreement and would go through a systematic moderation process:

Quantify the disagreement

o See how many bands (for example 1 or 3) big the gap is.

o Consider whether the disagreement is to all the score or just some of it (e.g., 'coherence' and not 'grammar').

Find the Cite within your Rubric

o Re-visit the description of the criterion in question.

o Check for vagueness ("good", "clear", "appropriate" with no operational definitions).

o Validate concentration levels don't overlap (i.e., Level 3 and Level 4 pertain the same performance).

o If there is ambiguity/overlap, the rubric will likely be a source.

Check rater behavior (rater effects)

o Analyze each rater's scoring for multiple scripts/samples.

o If one rater tends to be routinely harder or easier, that indicates severity/leniency bias.

o If the same rater is not consistent between Samples, it indicates poor intra-rater reliability (fatigue, changing standards, dissatisfaction with calibration).

Review rater training and calibration

o Ask: Did raters receive training on anchor samples? Did they talk about what the individual bands look like?

o Use benchmark scripts / recordings [2] and score them with both raters, finally compare these scores with an established "gold standard".

o If the match becomes stronger for some, than calibration was rater interpretation not rubric.

Evaluate the task itself

o Does the task permit multiple valid interpretations (i.e., vague prompts)?

o Diagnose construct-irrelevant source (familiarity with the topic, imbalance in group work forces) memorized speech.

o If the evidence produced is uneven (some students can "perform well" but not show mastery of a critical skill), it's the task that needs to be addressed.

Moderation decision and action

o Use discussion in a reasonable manner (as necessary and appropriate) of the score tits to the benchmarks, if any, to arrive at a final score.

o Then decide the fix:

Revise descriptors if rubric is not clear,

- retrain raters if interpretation differs,

- redeveloping the task if it is not capturing or tapping into the desired construct.

Conclusion: There is a way of being professional which means you combine evidence (patterns across scripts) + calibration against benchmarks and task analysis rather than just subjective negotiation.

3) Assessment of product or process: which is important?

Question: What is the distinction between evaluating the product (the final presentation/essay) and the process (drafting/rehearsal)? Which one should I select and why?

Master correction (model answer)

Assessment is also end-product oriented (final essay, final presentation). It is an objective observation on what has been achieved at a particular point in time and it can be helpful when

making summative judgments such as grading or certification, checking that learning outcomes have been reached. It is generally easier to score and standardize product assessment, leading to better comparability.

The process reveals how learning occurs: planning, drafting, peer feedback, revisions, rehearsal, self-reflection and strategy use. Why is process assessment (as we're describing it) so closely related to formative assessment? Because both are developmental, can be used supportively for individual autonomy and produce feedback that can actually be used. It is particularly helpful in learning language because it reinforces actions that result in progress (e.g., rewriting for coherence, taking onboard feedback, practicing pronunciation).

It depends on what the purpose is:

• If the aim is accountability/certification then product evaluation is usually more suitable because it assesses end-of-course achievement.

• If the goal is learning and progress, then process evaluation has more value since it reflects improvement and encourages skills development.

Two examples of Master-level didactics where such a stance is arguable are:

• Product evaluation is the way of quality as well as success.

• Process evaluation safeguards learning quality, and prevents unfairness (weak final product with strong progress; strong one with strong outside help).

Conclusions: Process assessment is more usually meaningful for assessment for learning, whereas product assessment is indispensable for assessment of learning. A hybrid approach that drawn on these two mechanisms might provide the best model of impact (e.g. 70% product + 30% process with explicit criteria and evidence) strongest design: a combination of both.