

5.3 Evaluate an argument's relevance, measurement and reliability evidence

Name: _____ Class: _____ Date: _____

Total: 24 marks

KEY WORDS validity 效度 • reliability 信度 • bias 偏差

Objective

Evaluate an argument's relevance, measurement and reliability evidence. This is original GAC014-aligned practice, not an official assessment. All people, texts, records and instructions are fictional. This sheet's word ranges and analytic points are local teaching conventions, not official assessment specifications. Written scripts and plans support listening/speaking preparation; they do not replace reviewed audio or assess pronunciation, actual delivery or one-pass listening comprehension.

1 Key knowledge and worked examples

■ **Key knowledge: repetition cannot be assumed and a source label is not a guarantee**

In evaluating an argument, ask whether its reasons support its conclusion. Measurement **validity** 效度 asks whether the evidence measures the intended thing. **Reliability** 信度 concerns consistency under suitable repeated conditions; one survey does not prove it.

Bias 偏差 can arise from selection, leading wording or incomplete reporting. Identify the possible mechanism rather than declaring a source false because you dislike its speaker. This invented rule-review scenario concerns academic reasoning, not actual law or legal rights.

■ **Source A: text or task brief**

Original written library-procedure discussion: The stated aim is to understand all 500 students' views about an invented late-return review procedure. A link on the library website received 20 voluntary responses. The question was, "How helpful is our fair new review procedure?" Sixteen selected helpful and four selected not helpful. No responses were collected offline, and no repeat survey or wording comparison was conducted. The record gives no information about the views of the other 480 students or actual outcomes of disputed cases.

The administrator reports these numbers and proposes a broader, neutrally worded review. This is a classroom scenario, not a statement of real legal requirements.

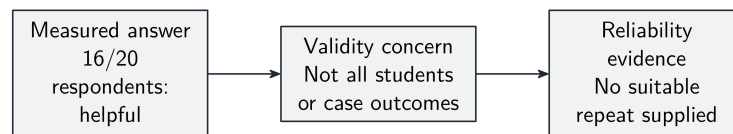
■ Source B: response or review record

Argument A: “The survey shows 80% of all students think every disputed case is handled fairly. Because it is on an official-looking page, the result is reliable and unbiased.”

Argument B: “The wording and selection may affect the answers, but that means every response is false.”

Evaluation task: assess what the count supports, the possible bias and what remains unknown about reliability. Do not invent responses or actual law.

■ Worked example: the measured answer is narrower than the proposed conclusion



Known: sixteen of twenty voluntary website respondents selected helpful. Why separate helpful from fair outcomes? The question did not measure what happened in every disputed case.

The words fair and helpful may encourage a positive framing, and website-only self-selection may omit other views. These are possible mechanisms, not proof that every answer is false. With no repeat or comparable measurement, consistency has not been established.

2 Practice

2.1 State the number selecting helpful and the response denominator. [2]

2.2 State TWO things the survey did not directly measure. [2]

2.3 Identify THREE problems with Argument A.

[3]

2.4 Explain one selection-bias concern, one wording concern and why reliability remains unestablished.

[3]

3 Exam-style questions

3.1 Propose a stronger evaluation and explain why each of THREE changes addresses a specific concern.

[6]

Solutions

2.1

Sixteen selected helpful; twenty voluntary website responses.

2.2

Any two, B1 each: all 500 students' views; actual fairness/outcomes of every disputed case; consistency across repeated surveys.

2.3

Voluntary 20 responses do not establish 80% of all 500; helpful answer does not prove fairness of every case; official-looking page does not establish consistency or absence of bias.

2.4

Website-only voluntary responses may miss other views; fair/helpful wording frames the procedure positively; no suitable repeat/comparison supplies consistency evidence.

3.1

Broader accessible recruitment addresses missing/nonresponding views without guaranteeing representation; neutral wording measuring the intended judgement addresses leading framing/construct mismatch; a suitable repeated or comparable measurement supplies evidence about consistency rather than assuming it.

3.2

Local analytic rubric, $8 \times B1$: accurate sixteen-of-twenty result; respondent population versus all five hundred; helpfulness versus case fairness; possible wording and selection mechanisms; no assumed reliability without suitable repetition; does not label every response false; specific improvement proposals with remaining limitations; coherent truthful 80–110 word response. Example: "Sixteen of twenty voluntary website respondents selected helpful, but these answers do not establish the views of all five hundred students or the fairness of every disputed case. The positive wording may frame the answers, and website-only recruitment may miss other views. These concerns do not make every response false. An official-looking page also cannot establish reliability; no suitable repeated measurement is supplied. A stronger review would recruit more broadly through accessible routes, use neutral questions matching the intended judgement, and examine consistency under comparable conditions. It should report remaining limitations rather than present broader recruitment as a guaranteed cure."