7.13 Reliability and Validity in Psychological Research

AQA 7182 · about 60 minutes · Write-in edition

Total Marks

Name: _____ Date: ____ / ____ / _____

Answer ALL questions.

1 Define the term 'validity' in psychological research and name two different forms of validity that researchers check for.

(Total for Question 1 is 2 marks)

2 Explain briefly how test-retest reliability is assessed and give one practical way a psychologist could improve test-retest reliability for a self-report questionnaire about stress.

(Total for Question 2 is 3 marks)

3 Outline what is meant by internal validity in the context of an experiment testing whether background music affects concentration on a cognitive task. Mention one example of a confounding variable in this study.

(Total for Question 3 is 4 marks)

4 A laboratory study measures children's recall of a short story immediately after hearing it and then again one year later, to test memory retention. Identify one threat to internal validity in this design and suggest one solution to reduce that threat.

(Total for Question 4 is 3 marks)

5 A researcher is worried that social desirability bias is reducing construct validity of a face-to-face interview measuring antisocial behaviour. Outline three procedural changes that could be made to reduce social desirability bias, and explain why each would help.

(Total for Question 5 is 6 marks)

6 Explain how a double-blind procedure can improve internal validity in a drug trial, and give two practical drawbacks of implementing a double-blind design in psychological research.

(Total for Question 6 is 6 marks)

7 A field observation study tests whether morning commuters are more likely to show impatient behaviour on trains during a particular month. Discuss one reliability issue and one validity issue specific to field observations in this context, and suggest one method to improve each issue.

(Total for Question 7 is 6 marks)

8 Discuss ways in which psychologists can ensure their research is reliable and valid. In your answer, refer to both quantitative and qualitative methods, give examples of at least two specific procedures for improving reliability and at least two procedures for improving validity, and evaluate the strengths and limitations of these procedures. You should cover issues such as standardisation, inter-observer reliability, test-retest methods, face/concurrent/internal/ecological/temporal validity, and experimental controls where relevant.

(Total for Question 8 is 16 marks)

Mark scheme · 7.13 Reliability and Validity in Psychological Research

Question 1

definition: validity is the extent to which a test or method measures what it is supposed to measure oe **B1**

two forms named, e.g. internal validity and ecological validity (or face validity and concurrent validity) oe **B1**

Answer: Validity is how well a test or method measures what it is intended to measure. Examples of forms: internal validity and ecological validity.

Question 2

explains assessment: the same test is given to the same participants on two occasions and the scores are correlated to assess consistency oe **M1**

gives a practical improvement, e.g. standardising instructions and testing interval, and ensuring the interval is appropriate to avoid real change in stress oe **B1**

explains why the improvement helps: consistent administration reduces extraneous variation that would lower the correlation oe **A1**

Answer: Test-retest reliability is assessed by giving the same test to the same people on two occasions and correlating the scores. To improve it for a stress questionnaire, standardise instructions and use an appropriate interval between tests so participants' real stress is unlikely to change; this reduces extraneous variation and increases consistency.

Question 3

outline: internal validity is the extent to which a study demonstrates a causal relationship between the IV and DV, with results not due to confounding variables oe **M1**

applies to the context: here it means whether any difference in concentration is due to background music rather than other factors oe **M1**

gives one confounding variable example, e.g. differences in participants' baseline concentration ability or recent caffeine intake oe **B1**

explains why it is a confound in context, e.g. if one condition contains more naturally attentive participants or more caffeinated participants, concentration scores may differ for that reason rather than due to music oe **A1**

Answer: Internal validity is the extent to which the experiment can show that background music caused any change in concentration rather than other factors. A confounding variable could be participants' baseline concentration ability or recent caffeine intake, which would affect scores independently of the music condition.

Question 4

identifies a threat: internal validity may be affected because the test and retest are a year apart and developmental changes or historical events could alter memory performance oe **M1**

suggests solution: use a shorter follow-up interval or include age-matched control groups tested at the same times to separate developmental effects from memory decay oe **B1**

explains how solution helps: shorter interval reduces influence of maturation/history, while control groups allow separation of cohort effects from the memory effect being studied oe **A1**

Answer: Threat: internal validity is threatened because developmental changes or life experiences over a year could change memory independently of the original encoding. Solution: use a shorter follow-up interval or include age-matched control groups tested at the same times; this reduces maturation and historical influences or allows them to be controlled for.

Question 5

suggests change one, e.g. use anonymous self-administered questionnaires rather than face-to-face interviews oe **M1**

explains why this helps: anonymity reduces pressure to present oneself favourably, increasing honesty and construct validity oe **A1**

suggests change two, e.g. use indirect questioning or the randomized response technique to make sensitive items less threatening oe **M1**

explains why this helps: indirect techniques reduce the risk of biased answers by protecting respondents and making disclosure less risky, improving validity oe **A1**

additional suggestion: use a self-completed computer-assisted method rather than an interviewer, or assure confidentiality in writing oe **M1**

explains why this helps: computer methods and written confidentiality assurances reduce interviewer presence effects and social desirability concerns, increasing honest responding oe **A1**

Answer: Change 1: use anonymous self-completed questionnaires rather than face-to-face interviews, because anonymity reduces the pressure to give socially desirable answers and increases construct validity. Change 2: use indirect methods such as randomized response or computer-assisted self-interviews so participants can answer sensitive questions without revealing personal admissions, which reduces bias and improves validity. Additional measures, such as written confidentiality assurances, also help by reducing fear of disclosure.

Question 6

explains how double-blind works: neither participants nor experimenters know which condition (drug or placebo) participants are in oe	**M1**
explains how this improves internal validity: it reduces demand characteristics and experimenter expectancy effects that could influence participants' behaviour or the measurement, so observed differences are more likely due to the drug itself oe	**A1**
identifies a practical drawback, e.g. increased logistical complexity and cost, or difficulty when side effects make blinding impossible oe	**M1**
explains why this is a drawback: complexity and cost can limit feasibility, and if side effects unblind participants or researchers the intended control is lost, reducing the benefit oe	**A1**
gives a second drawback or limitation, e.g. ethical concerns about withholding information or practical issues with emergency unblinding procedures oe	**M1**
explains effect: such ethical or safety procedures may require revealing condition allocation, which can compromise blinding and internal validity oe	**A1**

Answer: A double-blind procedure means neither participants nor researchers know who receives the drug or placebo. This reduces experimenter expectancy and demand characteristics, so differences are more likely to reflect the drug effect, improving internal validity. Drawbacks include greater logistical complexity and cost, and the possibility that side effects or safety procedures can unblind the study, reducing feasibility and potentially negating the benefits.

Question 7

identifies a reliability issue: observer inconsistency or low inter-observer reliability in coding impatient behaviours across observers or days oe	**M1**
explains why this matters: inconsistent coding reduces the trustworthiness of comparisons across times and sites oe	**A1**
suggests improvement for reliability: use clear operational definitions, train observers and conduct inter-observer agreement checks or pilot coding sessions oe	**M1**
explains how this helps: training and agreement checks increase consistency and make findings more reliable oe	**A1**
identifies a validity issue: observers may misinterpret behaviours or context, so ecological validity may be high but construct validity of 'impatience' may be poor oe	**M1**
suggests improvement for validity: combine observations with brief anonymous passenger surveys or behavioural coding of validated indicators, and triangulate data to ensure the measure captures the intended construct oe	**A1**

Answer: Reliability issue: observers may code impatient behaviours inconsistently. Improvement: provide clear operational definitions, train observers and perform inter-observer agreement checks to raise reliability. Validity issue: observed acts may not accurately reflect 'impatience' as a construct. Improvement: triangulate with brief anonymous passenger self-reports or use validated behavioural indicators so the measure better captures the intended construct.

Question 8

Level 4 Knowledge of methods to ensure reliability and validity is detailed and accurate. Discussion evaluates a good range of specific procedures for improving reliability and validity in both quantitative and qualitative contexts, showing balance and depth. Examples are applied appropriately, with implications and limitations considered. The answer is coherent, well organised and uses relevant terminology.	**13-16**
Level 3 Knowledge is clear with several relevant points about reliability and validity. Discussion includes evaluation and some development of procedures, with reasonable application to examples. Some balance between strengths and limitations is evident. Organisation is generally clear and terminology is used appropriately.	**9-12**
Level 2 Knowledge is present but partial, with limited range of procedures. Discussion is descriptive rather than evaluative, or evaluation is superficial. Examples or application may be brief or not well developed. The answer may lack organisation in places.	**5-8**
Level 1 Knowledge is minimal and answers are likely to be fragments or lists. Little or no evaluation is present. Application and examples are absent or irrelevant. Answer lacks structure and specialist terminology is mostly missing.	**1-4**
Level 0 No creditable content.	**0**

Indicative content:

- AO1: Definitions and distinction, reliability as consistency (test-retest, inter-observer, internal consistency) and validity as measuring what is intended (face, concurrent, construct, internal, ecological, temporal).
- AO1: Standardisation in quantitative experiments improves reliability by keeping procedures, instructions, timing and equipment identical, and improves internal validity by reducing procedural confounds.
- AO2: Example: in a reaction-time lab, standardising instructions and using the same apparatus increases test-retest reliability and reduces measurement error; but it may reduce ecological validity because the lab task is artificial.
- AO1: Inter-observer reliability procedures include operational definitions, observer training, pilot coding and calculating agreement coefficients (e.g. Cohen's kappa or percentage agreement).
- AO2: Example: classroom observation study can increase inter-observer reliability by trial coding sessions; this raises consistency but requires resources and may not eliminate shared bias.
- AO1: Test-retest reliability is assessed by correlating scores from two administrations; split-half and internal consistency methods (e.g. Cronbach's alpha) apply to questionnaires.
- AO3: Evaluation of reliability procedures: strengths include increased replicability and measurement precision; limitations include practice effects for test-retest, time and cost of training observers, and the possibility that too much standardisation removes naturalistic variation.
- AO1: Ways to improve validity include use of control groups and random allocation to strengthen internal validity, use of double-blind procedures to reduce expectancy effects, concurrent validity checks with established measures, and triangulation in qualitative research.
- AO2: Example: a new social anxiety scale can be validated by correlating with clinical interview scores (concurrent validity) and by testing predictive relationships with real-world outcomes, while using a diverse sample improves external validity.
- AO3: Evaluation of validity procedures: control groups and blinding increase confidence in causal claims but can be ethically or practically difficult; concurrent validity depends on the quality of the comparison measure; triangulation boosts construct validity in qualitative work but may produce conflicting data that are hard to reconcile.
- AO3: Consider trade-offs and practical constraints: improving internal validity through laboratory control may reduce ecological validity, and ensuring high reliability does not guarantee validity. Discuss temporal validity issues and methods to check it, such as replication across time and cultures.
- AO2/AO3: Conclude with balanced judgement: multiple complementary techniques are needed; researchers must choose procedures appropriate to their aims, and transparently report methods so others can judge reliability and validity.