



PE.PT8 Evaluating Fitness Test Results: Reliability, Validity and Normative Data

AQA 8582 · Calculators not allowed · about 40 minutes

Total Marks

--

Name: _____ Date: ____ / ____ / ____

Answer ALL questions. Show all your working.

- 1** In the context of fitness testing for a school PE class, tick one box to identify the term meaning that a test gives very similar results when repeated under the same conditions.
- ☐ A) Validity
 - ☐ B) Reliability
 - ☐ C) Objectivity
 - ☐ D) Practicality

(Total for Question 1 is 1 mark)

- 2** State the meaning of 'validity' in fitness testing as it applies to deciding whether a test is appropriate for a particular sport, for example whether a multi-stage fitness test measures a swimmer's endurance.

(Total for Question 2 is 1 mark)

- 3** Identify one practical reason a school might choose a fitness test that is less valid but more practical than an ideal laboratory test.

(Total for Question 3 is 1 mark)

- 4** State one factor that could reduce the reliability of a 30 m sprint test when the same athlete repeats the test on two different days.

(Total for Question 4 is 1 mark)

- 5** Table A shows normative data for male 15 year olds for three common tests. Use the table and read off: a 30 m sprint time of 4.35 seconds, a vertical jump of 44 cm, and a sit-and-reach of 28 cm. For each score, state the performance rating band from Table A.

Table A, male age 15 normative bands

30 m sprint time: Excellent ≤ 4.20 s, Good 4.21-4.50 s, Average 4.51-4.80 s, Below average > 4.80 s

Vertical jump height: Excellent ≥ 50 cm, Good 45-49 cm, Average 40-44 cm, Below average < 40 cm

Sit-and-reach flexibility: Excellent ≥ 35 cm, Good 30-34 cm, Average 25-29 cm, Below average < 25 cm

- (a) For the 30 m sprint time of 4.35 seconds, state the rating from Table A for a male aged 15. (1)
- (b) For the vertical jump height of 44 cm, state the rating from Table A for a male aged 15. (1)
- (c) For the sit-and-reach score of 28 cm, state the rating from Table A for a male aged 15. (1)

(Total for Question 5 is 3 marks)

-
- 6** A teacher uses hand timing with a stopwatch for a 30 m sprint. State one weakness of hand timing compared with electronic timing that affects reliability.

(Total for Question 6 is 1 mark)

-
- 7** Explain why a field-based test such as the sit-and-reach may be valid for checking a gymnast's flexibility but not valid for measuring a footballer's overall agility.

(Total for Question 7 is 1 mark)

- 8** State one way a tester could improve the reliability of vertical jump testing in a school setting.

(Total for Question 8 is 1 mark)

- 9** A coach compares a swimmer's multi-stage fitness test result with pool-based laps. State one reason why comparing a runner's shuttle run result with a swimmer's lap time may give invalid conclusions.

(Total for Question 9 is 1 mark)

- 10** Table B shows normative bands for female aged 17. Use Table B to rate two scores: a 30 m sprint time of 5.20 seconds and a sit-and-reach score of 33 cm.

Table B, female age 17 normative bands

30 m sprint time: Excellent ≤ 4.80 s, Good 4.81-5.10 s, Average 5.11-5.40 s, Below average > 5.40 s

Sit-and-reach flexibility: Excellent ≥ 36 cm, Good 31-35 cm, Average 26-30 cm, Below average < 26 cm

- (a) For the 30 m sprint time of 5.20 seconds, state the rating from Table B for a female aged 17. (1)
- (b) For the sit-and-reach score of 33 cm, state the rating from Table B for a female aged 17. (1)

(Total for Question 10 is 2 marks)

- 11** A gym uses a single tester to score sit-and-reach results for a group of 20 pupils. State one advantage and one disadvantage of using a single tester in terms of objectivity and reliability.

(Total for Question 11 is 2 marks)

- 12** State one reason why normative data should be age- and gender-banded when interpreting fitness test results.

(Total for Question 12 is 1 mark)

- 13** A tester records a vertical jump of 38 cm for a 16 year old male and then re-tests 10 minutes later and records 41 cm. State two possible explanations for this difference that relate to test quality.

(Total for Question 13 is 2 marks)

- 14** State one practical consideration a PE teacher must take into account when choosing a fitness test to use with a whole year group.

(Total for Question 14 is 1 mark)

- 15** Evaluate the reliability, validity and practicality of using the 30 m sprint test to assess a youth rugby player aged 14 who plays as a winger. Make a reasoned judgement about whether the 30 m sprint is an appropriate test for this player and suggest one improvement or alternative test. Use the rugby winger context in your answer.

(Total for Question 15 is 6 marks)

Mark scheme · PE.PT8 Evaluating Fitness Test Results: Reliability, Validity and Normative Data

Question 1

- **B1 B**
- **Answer: B (Reliability)**

Question 2

- **B1** it measures whether a test actually tests what it claims to measure, oe
- **Answer: Validity, it means the test actually measures what it claims to measure.**

Question 3

- **B1** examples such as: easier to set up / cheaper equipment / quicker to administer / can test many pupils at once, oe
- **Answer: For example, it is quicker to administer to a whole class and needs no expensive equipment.**

Question 4

- **B1** any one of: different wind conditions / different surface / inconsistent timing methods / fatigue or illness, oe
- **Answer: Different surface (one test on grass and the other on astroturf) which affects times.**

Question 5

- (a) **B1 Good**
- (a) **Answer: Good**
- (b) **B1 Average**
- (b) **Answer: Average**
- (c) **B1 Average**
- (c) **Answer: Average**

Question 6

- **B1** human reaction time introduces variable start/stop delays, reducing consistency between trials, oe
- **Answer: Human reaction time when starting and stopping the stopwatch introduces variable delays, reducing consistency.**

Question 7

- **B1** sit-and-reach measures hamstring and lower back flexibility, which is relevant to gymnastics but does not measure multi-directional speed and control needed for agility, oe
- **Answer: The sit-and-reach measures hamstring and lower back flexibility which is relevant to gymnastics, but it does not assess multi-directional speed or coordination required for football agility.**

Question 8

- **B1** examples such as: use the same warm up, use the same surface and shoes, conduct all trials at the same time of day, use a standardised measurement device, oe
- **Answer: Use the same warm up and test surface for every athlete and measure with the same device each time.**

Question 9

- **B1** they measure endurance under different modes and energy demands; running is weight-bearing and uses different muscle groups and movement patterns than swimming, so results are not directly comparable, oe
- **Answer: They use different movement patterns and muscle groups, so the tests measure endurance in different modes and are not directly comparable.**

Question 10

- (a) **B1 Average**
- (a) **Answer: Average**
- (b) **B1 Good**
- (b) **Answer: Good**

Question 11

- **B1** advantage: reduces inter-tester variation so results are more consistent between pupils, improving reliability, oe
- **B1** disadvantage: if the tester is biased or inconsistent the same error affects all scores and objectivity may be reduced, oe
- **Answer: Advantage: using one tester reduces differences between testers, improving reliability. Disadvantage: if that tester is biased or inconsistent their error affects every score, harming objectivity.**

Question 12

- **B1** because physical capacities change with growth and hormonal differences, so comparing like-for-like (age and gender) gives fairer interpretation, oe
- **Answer: Because physical capacities change with growth and between genders, using age and gender bands allows fairer comparisons.**

Question 13

- **B1** first reason: inadequate warm up before the first trial, leading to lower performance, oe
- **B1** second reason: measurement inconsistency or tester error between trials, oe
- **Answer: Possible reasons: the athlete had an inadequate warm up before the first trial, and the measurement method or tester was inconsistent between trials.**

Question 14

- **B1** examples such as: available space/time, number of pupils, required equipment, safety, or the need for qualified supervision, oe
- **Answer: For example, available time and space to test the whole year group is a key practical consideration.**

Question 15

- **Level 3** (5-6): A well-developed evaluation that discusses reliability, validity and practicality in the context of a 14 year old rugby winger, uses evidence and examples to support points, compares strengths and limitations, and reaches a clear, justified judgement with an appropriate suggestion for improvement or an alternative test.
- **Level 2** (3-4): A reasonable evaluation that covers at least two of reliability, validity or practicality with some application to the rugby winger, offers a conclusion but with limited justification or a weak suggestion for improvement.
- **Level 1** (1-2): Simple statements about reliability, validity or practicality with little or no application to the rugby winger and no clear judgement.
- **Level 0** (0): No relevant content.
- **Indicative content:**
 - Reliability: the 30 m sprint can be reliable if conditions are standardised, using electronic timing or the same tester, same surface and warm up; variable surfaces, inconsistent timing or fatigue reduce reliability.
 - Validity: for a rugby winger the 30 m sprint measures straight-line speed which is relevant to sprinting down the wing, but it does not assess acceleration over the first 5-10 m, multi-directional speed or rugby-specific movements such as evasion, so it is partially valid rather than fully sport specific.
 - Practicality: the 30 m sprint is practical in most school or club settings, needs minimal equipment and time, and can test individuals quickly, but requires safe straight run area and some organisation for many players.
 - Judgement guidance: a balanced answer might conclude the 30 m sprint is an appropriate baseline test for speed for a 14 year old winger, but should be complemented by a short acceleration test (e.g. 5 or 10 m sprint) or an agility test to capture sport-specific demands; this improves validity while keeping practical benefits.
 - Improvement/alternative: suggest adding electronic timing to improve reliability, or use a 5 m and 30 m split test or a rugby-specific sprint with evasion cones to improve validity.