**PART 3**

**Read the text below and answer questions 27–40.**

**BEYOND STANDARDISED TESTING: THE RISE OF ALTERNATIVE ASSESSMENT**

A

Standardised testing has long served as the dominant instrument for comparing student performance across schools, districts, and nations. Its appeal lies in psychometric consistency: uniform questions, controlled timing, and scoring procedures that appear to minimise human subjectivity. Yet the same features generate sustained criticism. High-stakes tests can narrow curricula as teachers feel pressure to prioritise examinable content over broader inquiry. They can reward short-term coaching that improves test familiarity without strengthening durable understanding. Skills that are difficult to capture under timed, constrained conditions—extended reasoning, collaboration, revision, and communication—may be treated as secondary because they do not fit the format. As a result, many education systems have begun to ask whether the evidence produced by standardised tests is too thin for the decisions it is asked to support.

B

Alternative assessment has emerged partly as a response to this mismatch between what schools claim to value and what tests can measure. Advocates argue that authentic tasks—those resembling real disciplinary work—improve pedagogical validity by requiring students to demonstrate knowledge in use rather than knowledge in isolation. A history student, for example, may be asked to analyse sources and defend an interpretation; a science student may design an investigation and justify conclusions; a language learner may craft an extended argument for a specific audience. These approaches can generate richer diagnostic information for teachers and can motivate students by making learning goals intelligible. However, the same authenticity creates practical drawbacks. Tasks take longer to administer, require extended supervision, and often demand careful feedback cycles. In large systems, the time and labour involved can become a bottleneck, particularly when policymakers still expect comparable results at scale.

C

Equity is frequently invoked on both sides of the debate. Proponents argue that performance-based tasks can reduce certain forms of cultural bias by allowing multiple ways to demonstrate competence and by valuing explanation over guessing strategies. A student who struggles with speeded multiple-choice formats may show strong reasoning in an extended response or a practical demonstration. Critics counter that alternative assessment can reproduce socio-economic stratification in subtler ways. Projects may

reward access to technology, quiet workspaces, and adult guidance that are unevenly distributed across households. If the polished appearance of an outcome reflects unequal support more than learning, assessment becomes a mirror of privilege. For this reason, equity often depends on design constraints: clear boundaries on outside assistance, structured checkpoints, and sufficient in-class time so students can complete core work under comparable conditions rather than relying on home resources.

D

Within the broad category of alternative assessment, there is no single template; it is a family of formats that reveal capabilities hidden in conventional test items. Portfolios assemble work across a term or year, allowing students to show progression, revise in response to critique, and reflect on their learning strategies. Performance tasks ask learners to apply knowledge to realistic problems, often integrating reading, writing, quantitative reasoning, or practical skill. Project-based assessment requires sustained inquiry, where students investigate a question, select evidence, and produce a product, report, or presentation. Each format shifts emphasis from rapid recall toward interpretation, construction, and communication. It also changes what teachers can see. Instead of a single score, educators observe how students plan, make errors, incorporate feedback, and justify decisions—process evidence that can be educationally valuable even when final products vary.

E

The central technical challenge is comparability. Standardised tests achieve consistency through identical questions and scoring rules, whereas alternative assessment relies on human judgement, which can vary across teachers, schools, and communities. Reliability therefore becomes the crucial concern, especially when results influence progression or university entry. Systems attempt to reduce inconsistency through detailed rubrics, scorer training, and moderation meetings in which teachers calibrate expectations using shared samples. Some designs keep teachers as local scorers but add external reviews of selected work to check drift and signal accountability without removing professional insight. Yet achieving stable standards remains difficult when resources differ across schools. Where staff workloads are high and training is limited, rubric use can become superficial, and moderation can be rushed. In such contexts, alternative assessment may produce precisely the problem it aims to solve: outcomes that are less trusted because they are not perceived as comparable.

F

Digital platforms are accelerating experimentation while introducing new risks. Online tools can collect evidence of learning, track revisions, and support peer feedback at scale,

making it easier to document process as well as product. Some systems also employ AI-assisted scoring to speed evaluation of writing or short responses, particularly where human marking would be slow or costly. However, automation raises concerns about opacity and accountability. If a score is produced by opaque algorithms, students and teachers may be unable to understand why a response was judged weak or strong, undermining procedural fairness. Bias in training data can also shift standards in ways that disadvantage particular language patterns or cultural styles. Moreover, when students know that an automated system will read their work, incentives may shift toward “writing for the model” rather than communicating clearly to human readers. For high-stakes decisions, many educators therefore treat AI as support rather than replacement, pairing automated suggestions with human review.

G

Because assessment functions as a public signal, policy alignment often determines whether alternative approaches can expand beyond pilot programmes. Universities and employers frequently demand simple, comparable indicators, which pulls systems back toward standardised scores even when educators prefer richer evidence. Where alternatives grow, they tend to do so alongside institutional recognition, investment in teacher training, and protected time for moderation and scoring. Some jurisdictions adopt hybrid models that combine standardised testing with performance tasks, aiming to balance broad comparability with deeper demonstrations of learning. In practice, the debate becomes less “tests versus no tests” than a question of evidentiary design: which measures are fit for which purposes, and how transparent systems are about what their evidence can—and cannot—support. The long-term viability of alternative assessment may therefore depend on whether systems can scale quality without overwhelming teachers, and whether stakeholders accept that fairness includes not only identical measurement, but also meaningful representation of what students can do.

**Questions 27–32**

Choose the correct heading for each paragraph.

**List of Headings**

**i** Criticisms of standardised tests and the move towards broader evidence
**ii** A range of alternative formats that reveal skills hidden in multiple-choice tests
**iii** Authentic tasks and deeper learning—along with practical drawbacks
**iv** Methods used to reduce inconsistency in judgement-based scoring
**v** Equity as a double-edged issue in project and performance assessment
**vi** Automation as a complete substitute for human marking

**vii** Digital platforms, AI-assisted scoring, and new concerns
**viii** Why resource demands may limit large-scale adoption

**27** Paragraph **A**

**28** Paragraph **B**

**29** Paragraph **C**

**30** Paragraph **D**

**31** Paragraph **E**

**32** Paragraph **F**

**Questions 33–36**

Choose **TRUE** if the statement agrees with the information given in the text, choose **FALSE** if the statement contradicts the information, or choose **NOT GIVEN** if there is no information on this.

**33** The passage suggests that alternative assessment always reduces teacher workload.

**34** The passage suggests that rubrics and moderation meetings are used to reduce inconsistency in scoring.

**35** The passage provides evidence that some universities now accept portfolios or performance tasks as official admissions evidence.

**36** The passage suggests that alternative assessment can expand without institutional recognition or investment in teacher training.

**Questions 37–40**

Complete the notes below. Choose **NO MORE THAN TWO WORDS** from the text for each answer.

**Key issues for alternative assessment**

- When human judgement varies among scorers, the crucial concern is **37** ..... .
- Equity can be improved through clear limits on outside assistance and structured **38** ..... .
- Automated scoring raises questions about responsibility for decisions, a concern described as **39** ..... .

- Some **40** ..... adopt hybrid models combining standardised testing with performance tasks.

Answer Key

27. i
28. iii
29. v
30. ii
31. iv
32. vii
33. FALSE
34. TRUE
35. NOT GIVEN
36. FALSE
37. reliability
38. checkpoints
39. accountability
40. jurisdictions