# Psychology subject report

2024 cohort

January 2025

*For all Queensland schools*

250038

# Contents

# Introduction

The annual subject reports seek to identify strengths and opportunities for improvement of internal and external assessment processes for all Queensland schools. The 2024 subject report is the culmination of the partnership between schools and the QCAA. It addresses school-based assessment design and judgments, and student responses to external assessment for General and General (Extension) subjects. In acknowledging effective practices and areas for refinement, it offers schools timely and evidence-based guidance to further develop student learning and assessment experiences for 2025.

The report also includes information about:

- how schools have applied syllabus objectives in the design and marking of internal assessments
- how syllabus objectives have been applied in the marking of external assessments
- patterns of student achievement.

The report promotes continuous improvement by:

- identifying effective practices in the design and marking of valid, accessible and reliable assessments
- recommending where and how to enhance the design and marking of valid, accessible and reliable assessment instruments
- providing examples that demonstrate best practice.

Schools are encouraged to reflect on the effective practices identified for each assessment, consider the recommendations to strengthen assessment design and explore the authentic student work samples provided.

## Audience and use

This report should be read by school leaders, subject leaders, and teachers to:

- inform teaching and learning and assessment preparation
- assist in assessment design practice
- assist in making assessment decisions
- help prepare students for internal and external assessment.

The report is publicly available to promote transparency and accountability. Students, parents, community members and other education stakeholders can use it to learn about the assessment practices and outcomes for senior subjects.

## Subject highlights

**218**
schools offered Psychology

**5.49%**
increase in enrolment since 2023

**85.19%**
agreement with provisional marks for IA2 and IA3

# Subject data summary
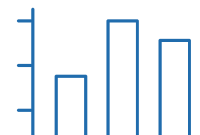
## Subject completion

The following data includes students who completed the General subject.

**Note:** All data is correct as at January 2025. Where percentages are provided, these are rounded to two decimal places and, therefore, may not add up to 100%.

Number of schools that offered Psychology: 218.

| Completion of units | Unit 1 | Unit 2 | Units 3 and 4 |
|---|---|---|---|
| Number of students completed | 6,488 | 5,952 | 5,244 |

## Units 1 and 2 results

| Number of students | Satisfactory | Unsatisfactory |
|---|---|---|
| Unit 1 | 5,990 | 498 |
| Unit 2 | 5,555 | 397 |

## Units 3 and 4 internal assessment (IA) results

### Total marks for IA

# IA1 marks

**IA1 total**



**IA1 Criterion: Data test**
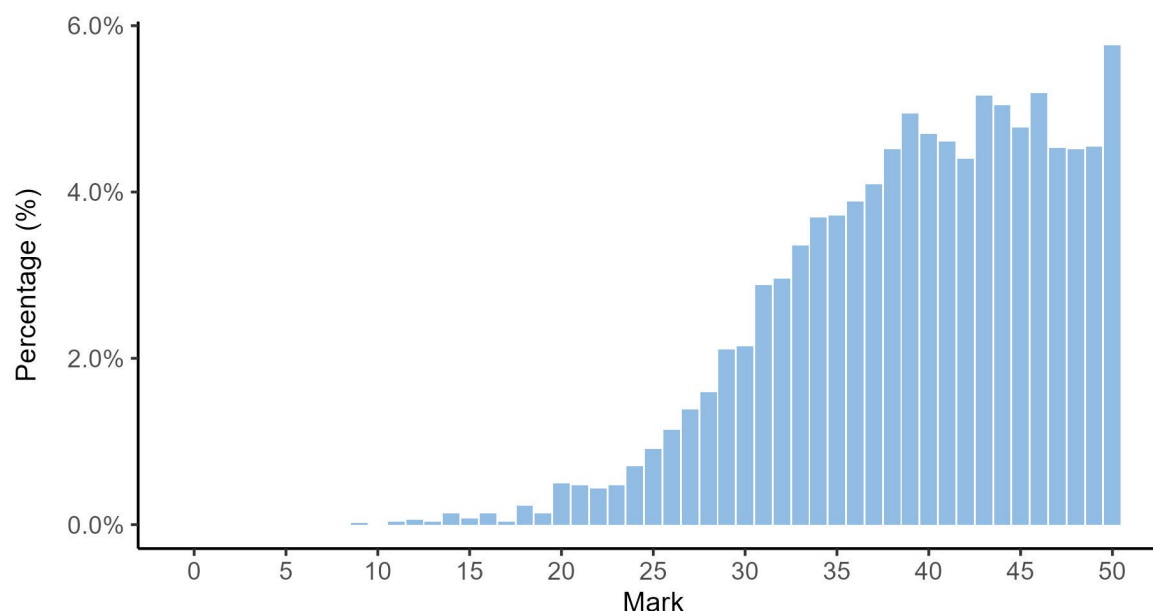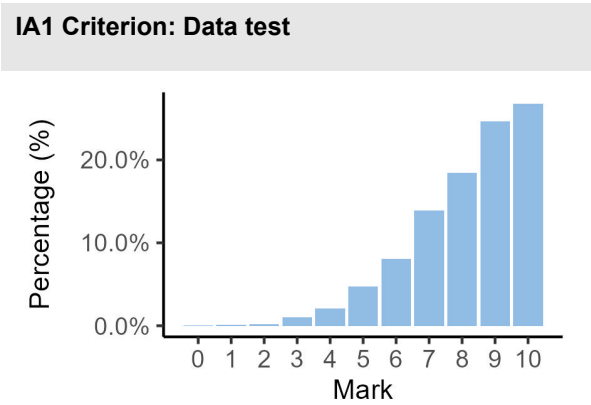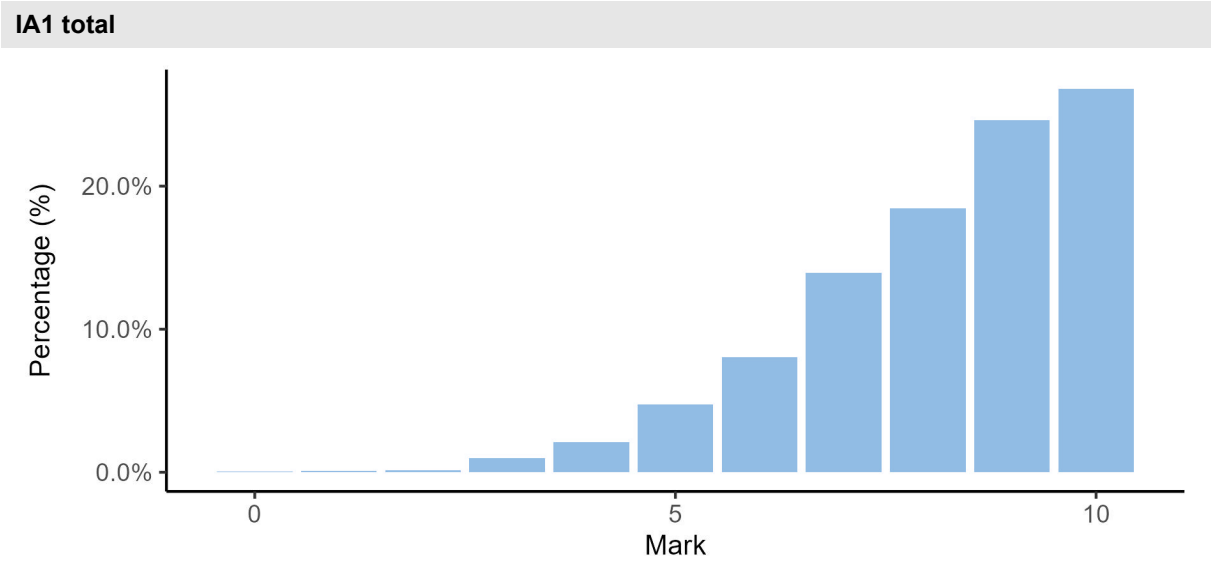
# IA2 marks

**IA2 total**
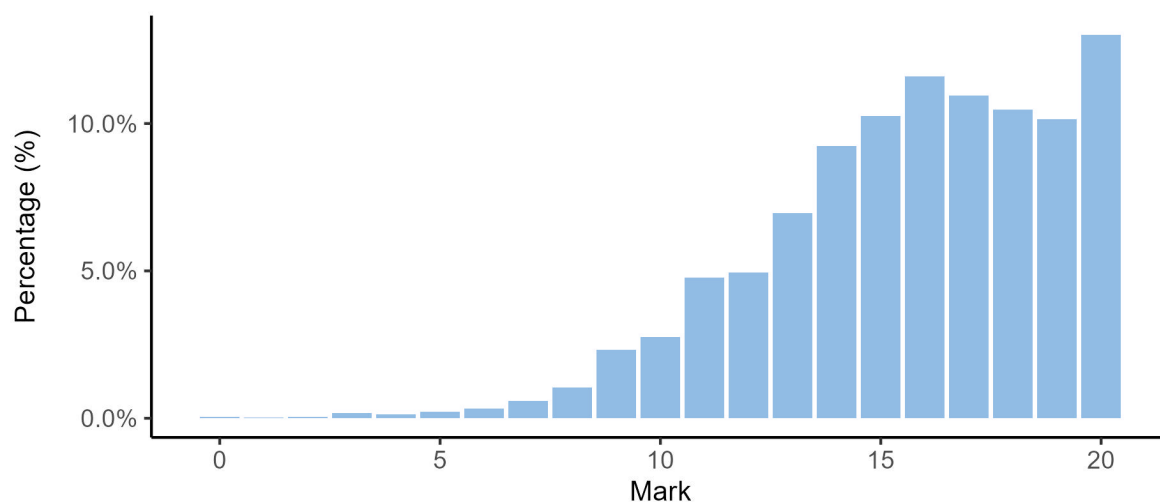


**IA2 Criterion: Research and planning**
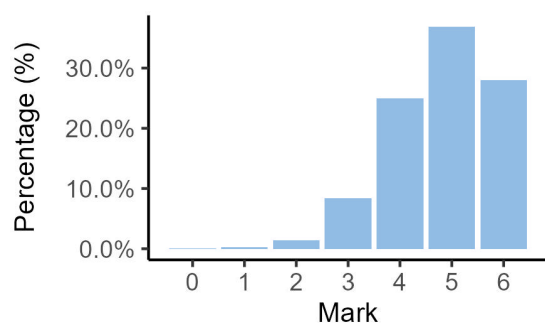


**IA2 Criterion: Analysis of evidence**
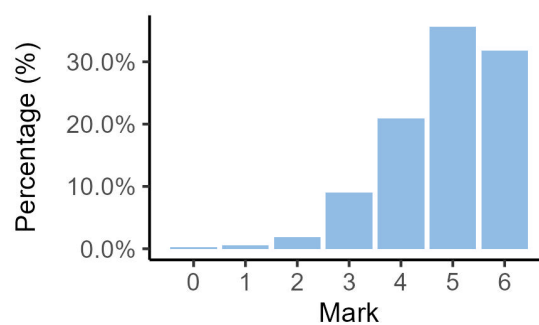


**IA2 Criterion: Interpretation and evaluation**
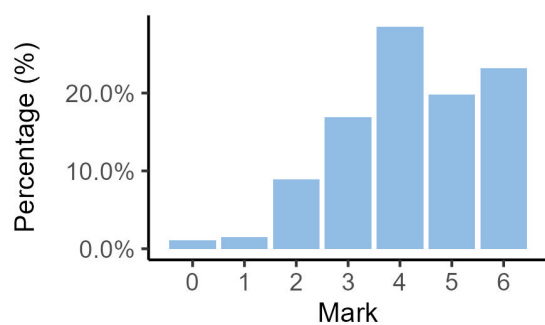


**IA2 Criterion: Communication**
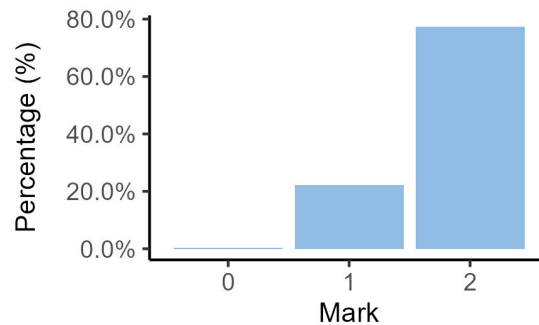
# IA3 marks

**IA3 total**
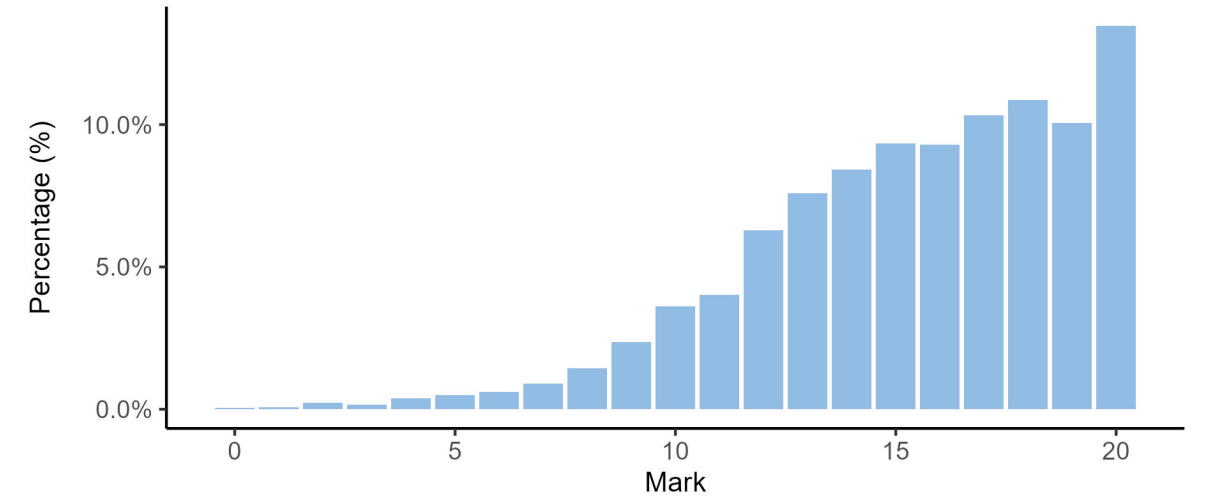


**IA3 Criterion: Research and planning**
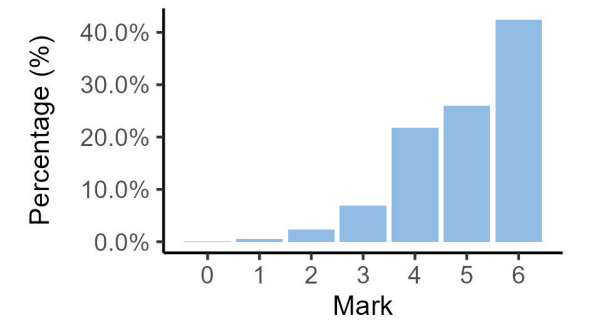


**IA3 Criterion: Analysis and interpretation**
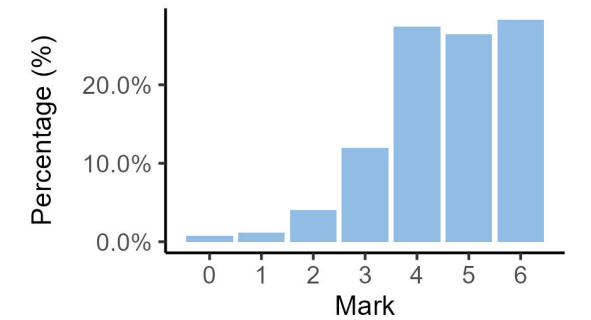


**IA3 Criterion: Conclusion and evaluation**
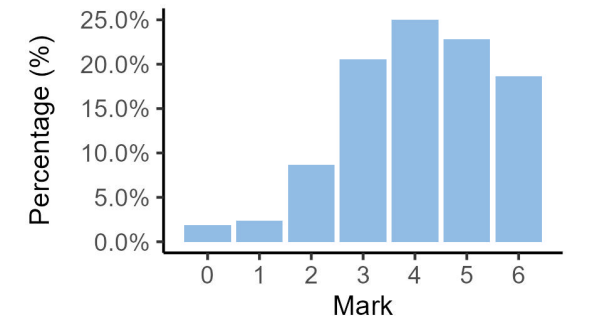


**IA3 Criterion: Communication**
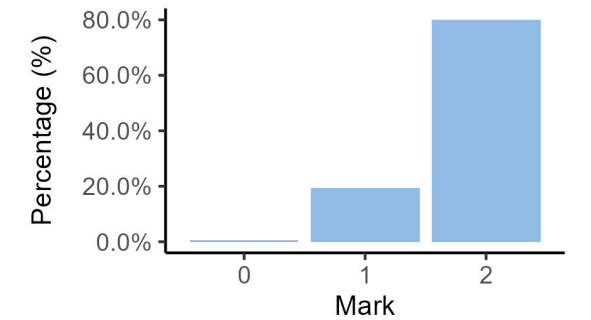
# External assessment (EA) marks
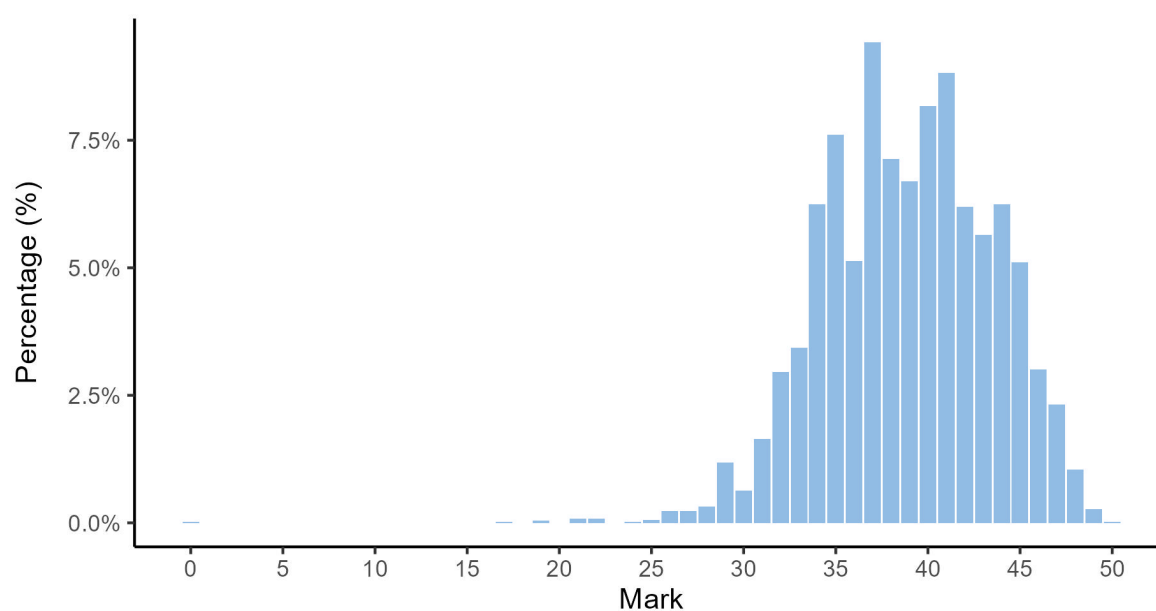


# Final subject results

## Final marks for IA and EA
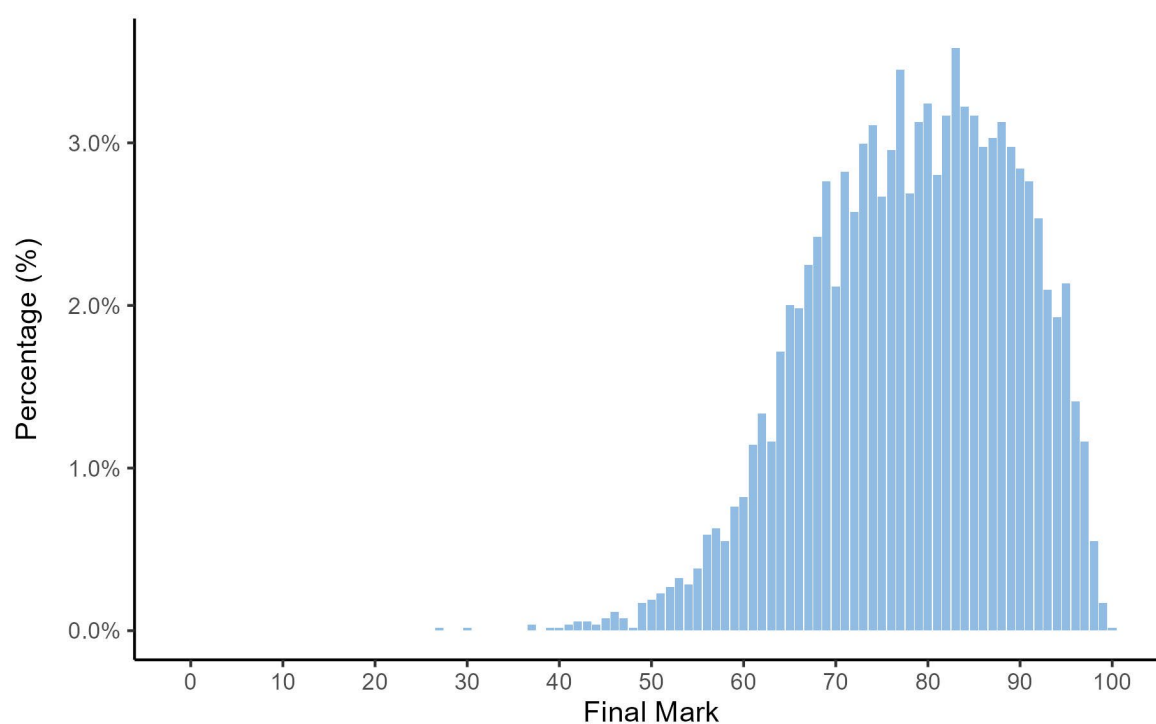
# Grade boundaries

The grade boundaries are determined using a process to compare results on a numeric scale to the reporting standards.

| Standard | A | B | C | D | E |
|---|---|---|---|---|---|
| **Marks achieved** | 100–87 | 86–72 | 71–49 | 48–20 | 19–0 |

# Distribution of standards

The number of students who achieved each standard across the state is as follows.

| Standard | A | B | C | D | E |
|---|---|---|---|---|---|
| **Number of students** | 1,403 | 2,398 | 1,412 | 31 | 0 |

# Internal assessment

The following information and advice relate to the assessment design and assessment decisions for each IA in Units 3 and 4. These instruments have undergone quality assurance processes informed by the attributes of quality assessment (validity, accessibility and reliability).

## Endorsement

Endorsement is the quality assurance process based on the attributes of validity and accessibility. These attributes are categorised further as priorities for assessment, and each priority can be further broken down into assessment practices.

Data presented in the Assessment design section identifies the reasons why IA instruments were not endorsed at Application 1, by the priority for assessment. An IA may have been identified more than once for a priority for assessment, e.g. it may have demonstrated a misalignment to both the subject matter and the assessment objective/s.

Refer to *QCE and QCIA policy and procedures handbook v6.0*, Section 9.6.

**Percentage of instruments endorsed in Application 1**

| Instruments submitted | IA1 | IA2 | IA3 |
|---|---|---|---|
| Total number of instruments | 218 | 218 | 216 |
| Percentage endorsed in Application 1 | 50 | 76 | 62 |

## Confirmation

Confirmation is the quality assurance process based on the attribute of reliability. The QCAA uses provisional criterion marks determined by teachers to identify the samples of student responses that schools are required to submit for confirmation.

Confirmation samples are representative of the school's decisions about the quality of student work in relation to the instrument-specific marking guide (ISMG), and are used to make decisions about the cohort's results.

Refer to *QCE and QCIA policy and procedures handbook v6.0*, Section 9.6.

The following table includes the percentage agreement between the provisional marks and confirmed marks by assessment instrument. The Assessment decisions section of this report for each assessment instrument identifies the agreement trends between provisional and confirmed marks by criterion.

**Number of samples reviewed and percentage agreement**

| IA | Number of schools | Number of samples requested | Number of additional samples requested | Percentage agreement with provisional marks |
|---|---|---|---|---|
| 1 | 216 | 1672 | 0 | 100.00 |
| 2 | 216 | 1680 | 31 | 85.19 |
| 3 | 216 | 1679 | 25 | 85.19 |

# Internal assessment 1 (IA1)

## Data test (10%)

This assessment focuses on the application of a range of cognitions to multiple provided items.

Student responses must be completed individually, under supervised conditions, and in a set timeframe.

## Assessment design

### Validity

Validity in assessment design considers the extent to which an assessment item accurately measures what it is intended to measure and that the evidence of student learning collected from an assessment can be legitimately used for the purpose specified in the syllabus.

**Reasons for non-endorsement by priority of assessment**

| Validity priority | Number of times priority was identified in decisions |
|---|---|
| Alignment | 83 |
| Authentication | 0 |
| Authenticity | 4 |
| Item construction | 7 |
| Scope and scale | 35 |

**Effective practices**

Validity priorities were effectively demonstrated in assessment instruments that:

- ensured the cognitions required by items aligned to the corresponding objective in the mark allocations table (Syllabus section 4.7.1), e.g. an item classified as Objective 2 (Apply understanding) should direct students to *determine* unknown scientific quantities or features
- contained unique datasets that were drawn from Unit 3 subject matter, and were reflective of actual studies
- included topics in the conditions that aligned to the datasets.

**Practices to strengthen**

It is recommended that assessment instruments:

- only include items that require analysis of the dataset and align to Objectives 2, 3 or 4. Items about methodology, hypotheses, or experimental design align to Objectives 5 and 6 and are not suitable for a data test
- have sufficient data for students to respond to the items, e.g. inferential statistics are provided when students are required to draw a conclusion
- avoid using datasets that are in the public domain to ensure that the stimulus is unseen.

## Accessibility

Accessibility in assessment design ensures that no student or group of students is disadvantaged in their capacity to access an assessment.

**Reasons for non-endorsement by priority of assessment**

| Accessibility priority | Number of times priority was identified in decisions |
|---|---|
| Bias avoidance | 9 |
| Language | 17 |
| Layout | 7 |
| Transparency | 30 |

### Effective practices

Accessibility priorities were effectively demonstrated in assessment instruments that:

- used clear wording so that students were cued to provide all components of the expected response

- were free of spelling, grammatical and referencing errors.

### Practices to strengthen

It is recommended that assessment instruments:

- advise students to show their working when a mark is awarded for calculation

- feature clearly labelled error bars (e.g. standard error or confidence intervals) where students must use them to draw conclusions about the significance of differences between means.

# Assessment decisions

## Reliability

Reliability is a judgment about the measurements of assessment. It refers to the extent to which the results of assessments are consistent, replicable and free from error.

**Agreement trends between provisional and confirmed marks**

| Criterion number | Criterion name | Percentage agreement with provisional | Percentage less than provisional | Percentage greater than provisional | Percentage both less and greater than provisional |
|---|---|---|---|---|---|
| 1 | Data test | 100 | 0 | 0 | 0 |

### Effective practices

Accuracy and consistency of the application of the ISMG for this IA was most effective when:

- the marking scheme showed the expected response and how each mark was awarded

- student responses were clearly annotated to show where each mark was awarded

- student responses aligned to the cognitive verb in the question.

**Practices to strengthen**

To further ensure accuracy and consistency of the application of the ISMG for this IA, it is recommended that:
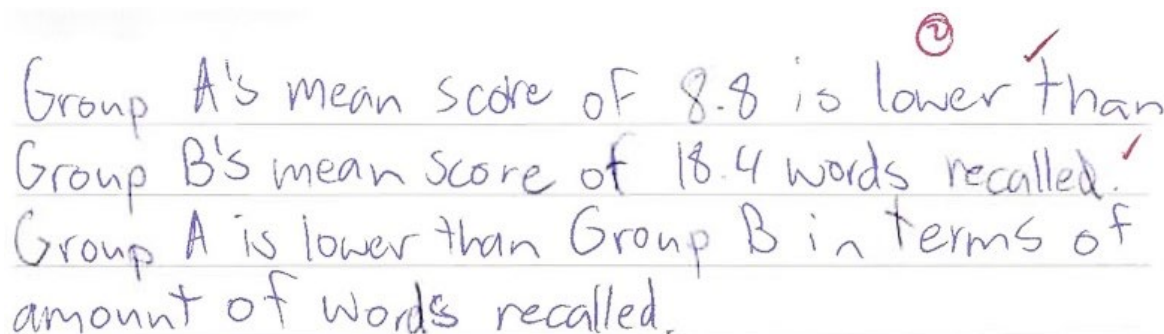
- marking schemes are updated to include alternative correct student responses and uploaded to the QCAA Portal Endorsement application (app)

- provisional marks are accurately calculated by correctly changing the mark to a percentage, then applying the percentage cut-offs on the ISMG e.g. 14/20 = 70%, which is >60%, so 7 should be awarded (see QCAA Portal > Syllabuses app > *Making judgments* webinar).

**Samples**

The following excerpts demonstrate clear annotation of student responses to indicate where marks have been awarded in relation to *contrast* (Excerpt 1) and *distinguish* (Excerpt 2) questions.
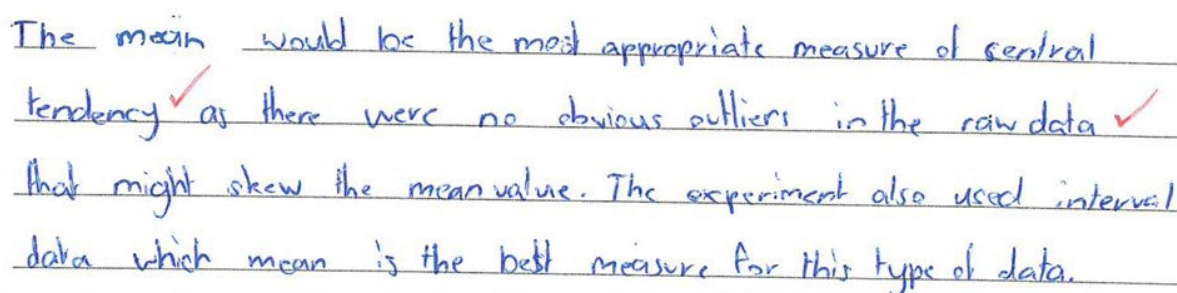
**Note:** The characteristics identified may not be the only time the characteristics occurred throughout a response.

**Excerpt 1**

Group A's mean score of 8.8 is lower than Group B's mean score of 18.4 words recalled. Group A is lower than Group B in terms of amount of words recalled.

**Excerpt 2**

The mean would be the most appropriate measure of central tendency as there were no obvious outliers in the raw data that might skew the mean value. The experiment also used interval data which mean is the best measure for this type of data.

## Additional advice

- Comparable assessments should be developed in the Endorsement app. They should be clearly labelled to ensure the correct assessment and its matching marking scheme are available for confirmation (*QCE and QCIA policy and procedures handbook v6.0*, Section 7.4).

# Internal assessment 2 (IA2)

## Student experiment (20%)

This assessment requires students to research a question or hypothesis through collection, analysis and synthesis of primary data. A student experiment uses investigative practices to assess a range of cognitions in a particular context. Investigative practices include locating and using information beyond students' own knowledge and the data they have been given.

Research conventions must be adhered to. This assessment occurs over an extended and defined period of time. Students may use class time and their own time to develop a response.

## Assessment design

### Validity

Validity in assessment design considers the extent to which an assessment item accurately measures what it is intended to measure and that the evidence of student learning collected from an assessment can be legitimately used for the purpose specified in the syllabus.

**Reasons for non-endorsement by priority of assessment**

| Validity priority | Number of times priority was identified in decisions |
|---|---|
| Alignment | 37 |
| Authentication | 6 |
| Authenticity | 0 |
| Item construction | 3 |
| Scope and scale | 6 |

**Effective practices**

Validity priorities were effectively demonstrated in assessment instruments that:

- clearly indicated an authentication strategy for comparing student work when working in groups

- contained a single draft submission

- identified practicals drawn from Unit 3 subject matter, i.e. Localisation of function in the brain, Visual perception, Memory, and/or Learning.

**Practices to strengthen**

It is recommended that assessment instruments:

- avoid using practicals that limit students' ability to demonstrate the range of performance levels in all objectives, e.g. Bugelski and Alampay (1961) produces categorical data and provides minimal scope for detailed analysis

- include only stimulus material relevant to completion of the task, and scaffolding that does not lead students to a predetermined response.

## Accessibility

Accessibility in assessment design ensures that no student or group of students is disadvantaged in their capacity to access an assessment.

### Reasons for non-endorsement by priority of assessment

| Accessibility priority | Number of times priority was identified in decisions |
|---|---|
| Bias avoidance | 0 |
| Language | 4 |
| Layout | 2 |
| Transparency | 3 |

### Effective practices

Accessibility priorities were effectively demonstrated in assessment instruments that:

- included clear checkpoints, e.g. each checkpoint on a new line with any calendar references clearly visible

- were free of grammatical and spelling errors.

### Practices to strengthen

There were no significant issues identified for improvement.

# Assessment decisions

## Reliability

Reliability is a judgment about the measurements of assessment. It refers to the extent to which the results of assessments are consistent, replicable and free from error.

### Agreement trends between provisional and confirmed marks

| Criterion number | Criterion name | Percentage agreement with provisional | Percentage less than provisional | Percentage greater than provisional | Percentage both less and greater than provisional |
|---|---|---|---|---|---|
| 1 | Research and planning | 94.44 | 5.09 | 0.46 | 0 |
| 2 | Analysis of evidence | 92.59 | 7.41 | 0.00 | 0 |
| 3 | Interpretation and evaluation | 93.06 | 6.94 | 0.00 | 0 |
| 4 | Communication | 99.07 | 0.00 | 0.93 | 0 |

### Effective practices

Accuracy and consistency of the application of the ISMG for this IA was most effective when:

- for the Research and planning criterion

  - a *considered* rationale clearly connected the research question to Unit 3 subject matter, using key terms and concepts

- the methodology clearly allowed the collection of *sufficient* and *relevant* data, as demonstrated by

    - a sufficiently high number of participants to allow identification of trends, patterns or relationships

    - data that allowed the research question to be answered

  - management of risks and ethical issues were *considered,* as demonstrated by the discussion of *how* and *why* a variety of risks and ethical issues were managed

- in the Communication criterion

  - findings, arguments and conclusions were *fluently* and *concisely* conveyed through precise and accurate use of

    - discipline-specific language

    - statistical language

    - indicators of uncertainty, e.g. measures of dispersion, and discussion of errors in hypothesis testing

    - tables, graphs and diagrams

  - *appropriate* use of genre conventions was demonstrated by the adherence to accepted rules of spelling and punctuation and the expectations of particular generic forms, e.g. for a scientific report — appropriate headings and captions, an appropriately formal tone, use of past tense

  - *appropriate* referencing conventions acknowledged sources through the consistent use of an accepted methodology, e.g. APA or Harvard.

## Practices to strengthen

To further ensure accuracy and consistency of the application of the ISMG for this IA, it is recommended that:

- in the Analysis of evidence criterion, *correct* and *relevant* processing of data includes

  - the most appropriate measurement of central tendency (mean or median)

  - alignment between measurements of dispersion and central tendency

  - the most appropriate graphical representation for the data, e.g. bar graph of mean values with error bars showing standard error of the mean or 95% confidence intervals

  - a suitable inferential test with a *p* value to test the null hypothesis

- in the Interpretation of evidence criterion

  - a *justified* conclusion answers the research question, using appropriate evidence, e.g. trends and patterns, statistical significance and/or reference to a *p* value

  - a *justified* discussion of reliability considers how the experimental design allowed the collection of consistent results with reference to the available evidence

  - a *justified* discussion of validity considers how the experimental design allowed the research question to be answered with reference to the available evidence. This may include population, ecological, or content validity.

## Samples

The following excerpts from a single student response demonstrate use of appropriate measures, graphical representations and use of inferential tests to support conclusions.

**Note:** The characteristic identified may not be the only time the characteristic occurred throughout a response.

**Excerpt 1**

Table 1: *Descriptive Statistics of number of correctly recalled answers out of 12 for both study conditions of participants aged 16-17.*

| Condition | n | Mean ($\bar{x}$) | (SD) | SEM | Skewness | Range |
|---|---|---|---|---|---|---|
| Silent Study Condition (SSC) | 21 | 7.00 | 2.32 | 0.51 | 0.21 | 9 |
| Noisy Study Condition (NSC) | 11 | 5.09 | 1.81 | 0.55 | 0.94 | 6 |

**Excerpt 2**
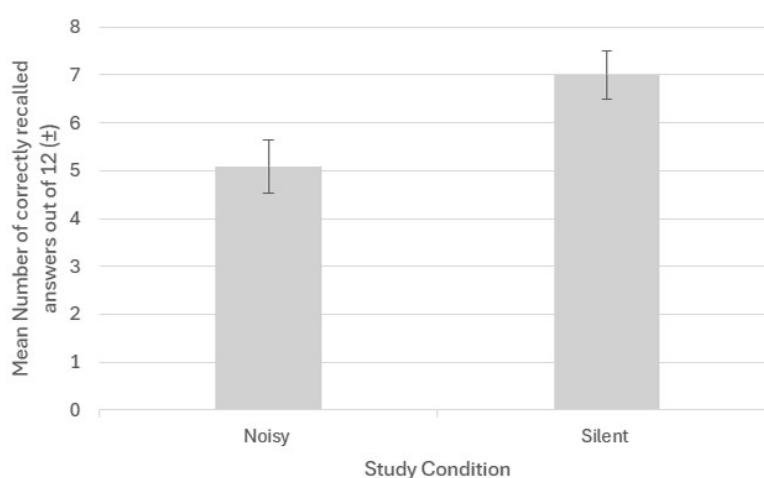


Figure 1: *Mean number (±) of correctly recalled answers (out of 12) in Noisy and Silent study conditions.*

**Excerpt 3**

For inferential statistics, a parametric independent t-test was selected to determine the statistical significance of the data (t=1.71, p=0.008). This type of t-test was conducted as the experiment design was independent groups with a directional alternative hypothesis. Since the p-value is less than the alpha level of 0.05, the data can be deemed statistically significant. Therefore, the null hypothesis is rejected, and the directional alternative hypothesis is accepted. Hence, context-dependency cues significantly impacted participants' ability to correctly recall short-response answers, supporting the research question.

The following excerpt demonstrates the use of appropriate evidence to justify a conclusion.

**Note:** The characteristic identified may not be the only time the characteristic occurred throughout a response.

---

### Conclusion

The research question states "Is there a statistically significant increase in chewing gum in matched study and testing conditions versus mismatched when Australian adolescent students recall study materials?" The p value is 0.48392 which means we will accept the null hypothesis. Based on the data there has been an extremely limited impact of gum seen as a context dependant cue to improve recall. This means in general the experiment shows that students who chew gum in study and testing have no positive effect.

---

## Additional advice

- Schools should ensure

    - academic integrity by including authentication strategies which monitor for plagiarism and manage authenticity concerns, including the use of any unattributed non-original material, through school-based processes (*QCE & QCIA policy and procedures handbook v6.0,* Section 8.1 and Section 8.2.8)

    - that strategies for responses that exceed syllabus conditions for length (2000 words) are developed and enacted. This includes annotating responses to show how the school's assessment policy has been applied (*QCE and QCIA policy and procedures handbook v6.0,* Section 8.2.6)

    - marked ISMGs indicate the characteristics evident in the student response and the mark awarded for each criterion (*QCE and QCIA policy and procedures handbook v6.0,* Section 9.6.1). Schools should consider evidence throughout the entire response when deciding which characteristics the evidence best matches

    - accuracy and consistency in judgments when determining grades for each criterion by applying the best-fit approach (see Syllabuses app > QCAA Portal > *Using ISMGs for General Science syllabuses*). After determining the performance level that best fits the evidence matched to characteristics for a criterion within an ISMG, for a two-mark range performance level, the higher mark should only be awarded if there is evidence of all the characteristics in the performance-level descriptor (or better).

# Internal assessment 3 (IA3)

## Research investigation (20%)

This assessment requires students to evaluate a claim. They will do this by researching, analysing and interpreting secondary evidence from scientific texts to form the basis for a justified conclusion about the claim. A research investigation uses research practices to assess a range of cognitions in a particular context. Research practices include locating and using information beyond students' own knowledge and the data they have been given.

Research conventions must be adhered to. This assessment occurs over an extended and defined period of time. Students may use class time and their own time to develop a response.

## Assessment design

### Validity

Validity in assessment design considers the extent to which an assessment item accurately measures what it is intended to measure and that the evidence of student learning collected from an assessment can be legitimately used for the purpose specified in the syllabus.

**Reasons for non-endorsement by priority of assessment**

| Validity priority | Number of times priority was identified in decisions |
|---|---|
| Alignment | 24 |
| Authentication | 6 |
| Authenticity | 1 |
| Item construction | 40 |
| Scope and scale | 17 |

**Effective practices**

Validity priorities were effectively demonstrated in assessment instruments that:

- clearly aligned claims to Unit 4 subject matter, suggested research or Science as a Human Endeavour (SHE) components
- ensured that there were relevant claims for each of the Unit 4 topics listed in the assessment conditions
- included authentication strategies referring to individual work only, as group work is not an aspect of IA3.

**Practices to strengthen**

It is recommended that assessment instruments:

- construct claims as single assertions that indicate a relationship between two broad variables, e.g. socialisation influences antisocial behaviour
- include claims that are within scope and scale of the task, being accessible at the level of the syllabus and addressable within the response length.

## Accessibility

Accessibility in assessment design ensures that no student or group of students is disadvantaged in their capacity to access an assessment.

### Reasons for non-endorsement by priority of assessment

| Accessibility priority | Number of times priority was identified in decisions |
|---|---|
| Bias avoidance | 0 |
| Language | 1 |
| Layout | 0 |
| Transparency | 1 |

### Effective practices

Accessibility priorities were effectively demonstrated in assessment instruments that:

- included claims that avoided cultural, religious and racial bias

- were free of grammatical and spelling errors.

### Practices to strengthen

There were no significant issues identified for improvement.

# Assessment decisions

## Reliability

Reliability is a judgment about the measurements of assessment. It refers to the extent to which the results of assessments are consistent, replicable and free from error.

### Agreement trends between provisional and confirmed marks

| Criterion number | Criterion name | Percentage agreement with provisional | Percentage less than provisional | Percentage greater than provisional | Percentage both less and greater than provisional |
|---|---|---|---|---|---|
| 1 | Research and planning | 97.22 | 1.85 | 0.93 | 0 |
| 2 | Analysis and interpretation | 96.30 | 3.70 | 0.00 | 0 |
| 3 | Conclusion and evaluation | 88.43 | 11.57 | 0.00 | 0 |
| 4 | Communication | 100.00 | 0.00 | 0.00 | 0 |

### Effective practices

Accuracy and consistency of the application of the ISMG for this IA was most effective when:

- in the Research and planning criterion, research questions were

  - clearly developed to be distinct from but still *relevant* to one aspect of the claim and to Unit 4 subject matter

- – *specific* in that they included one clear independent and one clear dependent variable
- in the Analysis and interpretation criterion
  - – *sufficient* and *relevant* evidence was selected from more than one credible source. The evidence was directly related to the research question so that relevant trends, patterns and relationships could be identified to form justified arguments about the question
  - – *thorough* and *appropriate* identification of limitations was specific to the research question and/or claim.

## Practices to strengthen

To further ensure accuracy and consistency of the application of the ISMG for this IA, it is recommended that:

- in the Conclusion and evaluation criterion
  - – conclusions directly answer the research question and are *justified* by credible findings from the research
  - – the quality of evidence is *insightfully* discussed, including discussion of the effect of limitations on the ability of the evidence to respond to the research question and the claim
  - – *credible* findings are specifically extrapolated to the claim, beyond simply addressing the research question
  - – suggested extensions explore aspects of the claim not addressed by the research question.

## Samples

The following excerpt demonstrates thorough identification of limitations in the generalisability of findings.

**Note:** The characteristic identified may not be the only time the characteristic occurred throughout a response.

---

**Limitations and Improvement**

The samples, while representative of their identified context, limit overall generalisability. Curşeu and Boroş (2008) focused on Romanian evaluators with tertiary education while Moss-Racusin et al. (2012) delved into science faculty across various fields. While they provide relevant information to their respective domain, extrapolating the findings is difficult without further studies. Culturally, gender roles and norms are subject to change based on the regions cultural assumptions surrounding gender. Thus, the gender schema instilled into a person's cognition is largely subject to change geographically and culturally. This indicates a new realm of study worth exploring in future research, to ascertain the difference in gender bias culturally.

Further, Curşeu and Boroş (2008) only assessed people with tertiary education, justifying it as ensuring participants could comprehend the task. However, not all promotional opportunities or people making these decisions possess this level of education. Building on this, Moss-Racusin et al. (2012) only assessed people within scientific fields. This proves highly useful in understanding the gender disparity in this male dominated field, however, does not indicate high levels of generalisability. To rectify limitations in both these studies, existing managers across workforce types should be surveyed to provide understandings based in real-world contexts. While both samples remained relevant to their contexts, their focus on specific cultures, education levels, or workplace types should be expanded to further understand the claim "gender determines behaviour."

---

The following excerpt demonstrates a conclusion that is justified by evidence from credible sources.

**Note:** The characteristic identified may not be the only time the characteristic occurred throughout a response.

---

This investigation aimed to answer, 'To what extent does the Hawthorne effect impact hand hygiene compliance among hospital workers?' under the claim 'The presence of others impacts how we behave', both of which have been supported to a large extent. Both Eduardo et al., 2022 ($p=0.016$, $p=0.048$) and Hagal et al., 2015 ($p<0.0001$) indicate that HH compliance increases under overt human observation because of the Hawthorne effect with statistical significance. However, Wu et al., 2018 ($p<0.001$ and $p=0.13$) indicates that HH compliance increases under human observation due to the Hawthorne effect with partial statistical significance. This investigation confirmed that HH compliance increases under human observation because the Hawthorne effect made participants behave differently to conform with societal expectations. However, additional research is required to further understand the implications and the Hawthorne effect.

---

The following excerpt demonstrates extrapolation of findings beyond the research question, and to the claim as a more general assertion.

**Note:** The characteristic identified may not be the only time the characteristic occurred throughout a response.

---

**Extrapolation to the Claim**

To relate this to the claim, it can be extrapolated that aggression is a learnt behaviour. Emotional regulation and one's ability to empathise with others is learnt from your parent's style, with negative correlations between warm parenting styles and emotional dysregulation (-0.27) and positive correlations between warm parenting and empathy levels (0.48). These factors directly affect aggressive behaviour, with a positive correlation between emotional dysregulation (0.52) and a negative correlation between empathy ability and aggression (-0.33). So therefore, if parenting styles teach emotional regulation and empathy, and these factors discourage aggression, it can be concluded that aggression is a learnt behaviour.

---

## Additional advice

- Schools should ensure

    - academic integrity by including authentication strategies which monitor for plagiarism and manage authenticity concerns, including the use of any unattributed non-original material, through school-based processes (*QCE & QCIA policy and procedures handbook v6.0,* Section 8.1 and Section 8.2.8)

    - that strategies for responses that exceed syllabus conditions for length (2000 words) are developed and enacted. This includes annotating responses to show how the school's assessment policy has been applied (*QCE and QCIA policy and procedures handbook v6.0,* Section 8.2.6)

    - marked ISMGs indicate the characteristics evident in the student response and the mark awarded for each criterion (*QCE and QCIA policy and procedures handbook v6.0,* Section 9.6.1). Schools should consider evidence throughout the entire response when deciding which characteristics the evidence best matches

        accuracy and consistency in judgments when determining grades for each criterion by applying the best-fit approach (see Syllabuses app > QCAA Portal > *Using ISMGs for General Science syllabuses*). After determining the performance level that best fits the evidence matched to characteristics for a criterion within an ISMG, for a two-mark range performance level, the higher mark should only be awarded if there is evidence of all the characteristics in the performance-level descriptor (or better).

# External assessment
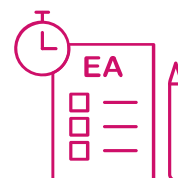
External assessment (EA) is developed and marked by the QCAA. The external assessment for a subject is common to all schools and administered under the same conditions, at the same time, on the same day.

## Examination (50%)

### Assessment design

The assessment instrument was designed using the specifications, conditions and assessment objectives described in the summative external assessment section of the syllabus.

The examination consisted of two papers:

- Paper 1, Section 1 consisted of multiple choice questions (20 marks)

- Paper 1, Section 2 consisted of short response questions (28 marks)

- Paper 2, Section 1 consisted of short response questions (44 marks)

### Assessment decisions

Assessment decisions are made by markers by matching student responses to the external assessment marking guide (EAMG). The external assessment papers and the EAMG are published in the year after they are administered.

#### Multiple choice question responses

There were 20 multiple choice questions in Paper 1.

**Percentage of student responses to each option**

**Note:**

- The correct answer is **bold** and in a blue shaded table cell.

- Some students may not have responded to every question.

| Question | A | B | C | D |
|---|---|---|---|---|
| 1 | 2.35 | 5.06 | **73.75** | 18.43 |
| 2 | 2.60 | **79.03** | 9.43 | 8.58 |
| 3 | **50.06** | 32.56 | 5.67 | 11.21 |
| 4 | 2.77 | 12.84 | 9.66 | **74.25** |
| 5 | **87.47** | 1.65 | 0.12 | 10.46 |
| 6 | 2.02 | 3.34 | **90.24** | 4.13 |
| 7 | 19.75 | 13.85 | **44.68** | 21.11 |
| 8 | 2.62 | **96.04** | 0.29 | 0.74 |
| 9 | **59.70** | 5.06 | 4.54 | 30.27 |
| 10 | 39.89 | 6.00 | 6.56 | **47.15** |

| Question | A | B | C | D |
|---|---|---|---|---|
| 11 | 9.39 | **34.73** | 31.10 | 24.39 |
| 12 | 6.11 | 1.01 | **92.20** | 0.37 |
| 13 | **71.40** | 14.42 | 1.79 | 11.99 |
| 14 | 2.74 | 3.43 | **68.28** | 25.26 |
| 15 | 7.86 | **78.09** | 5.59 | 8.15 |
| 16* | 31.14 | **15.41** | **30.83** | 22.22 |
| 17 | **50.19** | 18.72 | 18.84 | 11.82 |
| 18 | 33.64 | 10.42 | 30.44 | **24.99** |
| 19 | 8.01 | **63.14** | 8.36 | 20.12 |
| 20 | 15.93 | **36.03** | 18.65 | 29.01 |

*The multiple choice scrutiny panel reviewed the question and determined that there were two keys for this item.

## Effective practices

Overall, students responded well to short response questions that required:

- simple descriptions of key concepts using non-technical vocabulary

- identification of trends in graphed data relating to required or recommended readings in the syllabus

- conclusions to be drawn from tabulated data relating to required or recommended readings in the syllabus.

## Practices to strengthen

When preparing students for external assessment, it is recommended that teachers:

- provide opportunities for students to practice

  - describing concepts with precise use of technical vocabulary from subject matter e.g. classical and operant conditioning

  - expressing concepts unambiguously and accurately so that their responses can be matched to response characteristics in the external assessment marking guide

- provide opportunities for students to use familiar statistics relating to key research skills across a range of inquiry contexts

- spend time examining required readings with students so that they are familiar with the general methodology and key findings of these studies.

## Samples

### Short response

Paper 1, Question 24

The following excerpt is from Question 24 in Paper 1. It required students to describe evolutionary theories of attraction with reference to the finding of clear sex differences across cultures in Buss et. al. (1990).

Effective student responses:

- described evolutionary theories of attraction

- identified evidence of male preferences

- identified evidence of female preferences

This excerpt demonstrates accurate description of concepts being related to findings from required readings. It demonstrates knowledge of the findings, and effective use of terminology to provide an unambiguous response that aligned to the marking guide.

> Evolutionary theories of attraction suggest that mate preference is based on perceived and that sex impacts these preferences likelyihood of survival. for example, Buss et al.'s study found that females prefer males that show potential to earn lots which can be explained by the fact that it indicates that they will be able to provide for the female and possible offspring, increasing survival as they will have food, water, etc. chances. Males were found to prefer females that were physically attractive as this suggests fertility and therefore, a higher chance of their genes being passed to the next generation, this is also linked to increased survival

Paper 2, Question 2

The following excerpt is from Question 2 in Paper 2. It required students to use the concept of perceptual set to explain the findings in a scenario.

Effective student responses used the concept of perceptual set to explain the findings.

This excerpt demonstrates the accurate use of terminology to explain concepts and relate them to scenarios.

> The perceptual set is a set of psychological factors that impact a persons perception. One factor is someone's past experiences, which is that one's past experiences will impact how they percieve something. For the children who had no past experience with reading, but had past experience with learning colours, they responded more quickly because they only knew how to perceive the colours.
>
> Meanwhile, the children with past experience with reading responded lower, because they could perceive both the words and the colour, and had to spend more time figuring out which piece of information to use.

Paper 2, Question 3

The following excerpt is from Question 3 in Paper 2. It required students to refer to visual perception principles and the properties of the Ames Room to explain an illusion.

Effective student responses:

- described a relevant property of the Ames Room

- described an associated visual perception principle, and

- combined these to explain the illusion.

This excerpt demonstrates correct use of terminology about visual perception principles and specific knowledge of the Ames Room to explain the illusion.

> The Ames Room is a visual illusion in which shape constancy is used at the expense of size constancy. While the room is actually trapezoidal in shape, when viewed through a small hole, it appears square. The room must be viewed with only one eye so that binocular depth cues do not affect the illusion. Therefore when an individual walks to the other side of the room, they appear to shrink/grow when in actual fact they stay the same size.

Paper 2, Question 6d)

The following excerpts are from Question 6d in Paper 2. It required students to use data relevant to the Grant et. al. (1998) mandatory practical to determine confidence intervals and use them to draw a conclusion.

Effective student responses:

- used confidence intervals to determine the maximum value of the lower mean, and the minimum value of the higher mean

- determined that these values showed that the intervals did not overlap

- concluded that context-dependent cues affected memory

These excerpts demonstrate the use of statistical measures in a range of contexts. In both responses, tabulated data relevant to the Grant et al. (1998) mandatory practical was correctly used to determine confidence intervals to conclude that context-dependent cues affect memory.

**Excerpt 1**

Based on the data provided, ~~at~~ the confidence interval for the matched condition would range from 7.16 – 8.3, while the interval for the mismatched condition is 5.74 to 6.8. Since there is no overlap between these confidence intervals, it can be said that the results are statistically significant. This means that the ~~diff~~ difference in scores between the matched and mismatched conditions is likely due to matched/mismatched cues being used rather than chance.

**Excerpt 2**
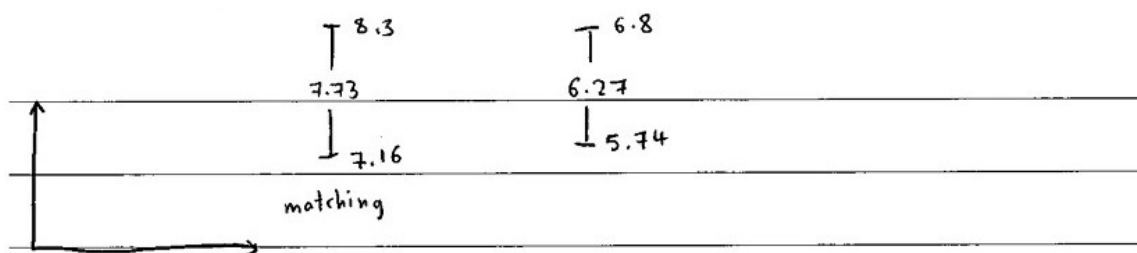


Based on the mean (7.73) and 95% confidence interval (0.57) of matching conditions, and mean (6.27) and 95% confidence interval (0.53) of mismatching conditions, as the confidence intervals do not overlap, there is a statiscally significant difference between the matching and mismatching conditions. Therefore, we reject the null hypothesis and the alternate hypothesis is supported. Based on these results, it ~~it~~ can be concluded that context-dependent cues affect memory and produce a higher mean score ~~than~~ (for matching rather than mismatching).