Psychology subject report

2025 cohort

January 2026

Queensland Government | QCAA Queensland Curriculum & Assessment Authority

260038

Queensland Curriculum & Assessment Authority
PO Box 307 Spring Hill QLD 4004 Australia

Phone: (07) 3864 0299
Email: office@qcaa.qld.edu.au
Website: www.qcaa.qld.edu.au

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 2025 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 2026.

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
- important considerations to note related to the revised 2025 syllabus (where relevant).

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

12.67%
increase in enrolment since 2024

85.65%
agreement with provisional marks for IA3

99.39%
of students received a C or higher

Subject data summary

Unit completion

The following data shows students who completed the General subject.

Note: All data is correct as at January 2026. 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: 237.

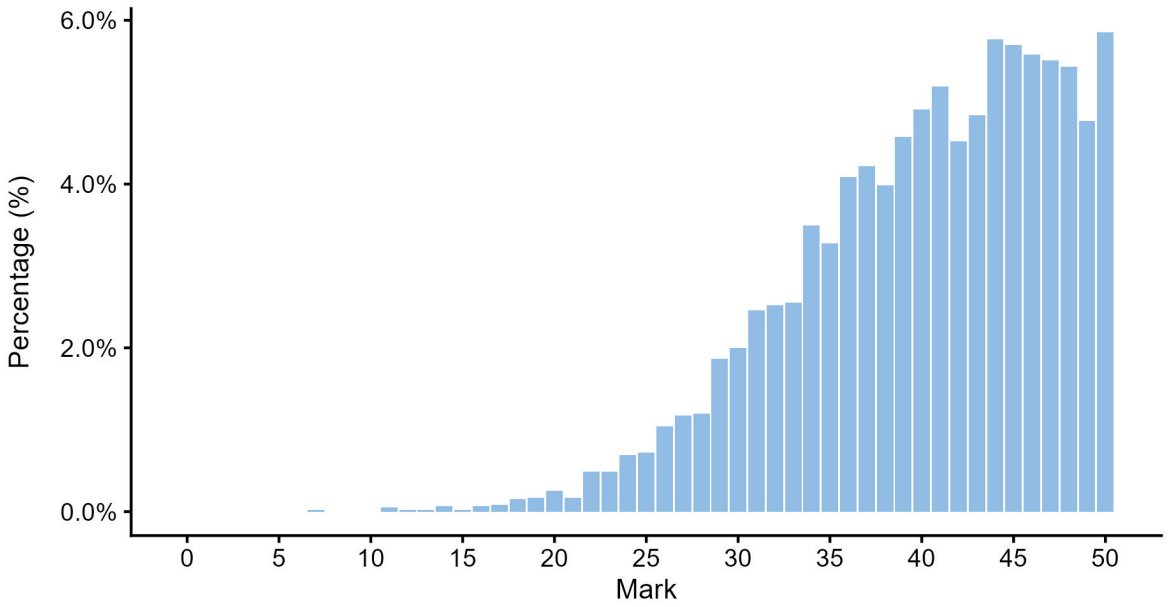
Completion of units	Unit 1	Unit 2	Units 3 and 4
Number of students completed	7,111	6,631	5.888

Units 1 and 2 results

Number of students	Unit 1	Unit 2
Satisfactory	6,591	6,208
Unsatisfactory	520	423

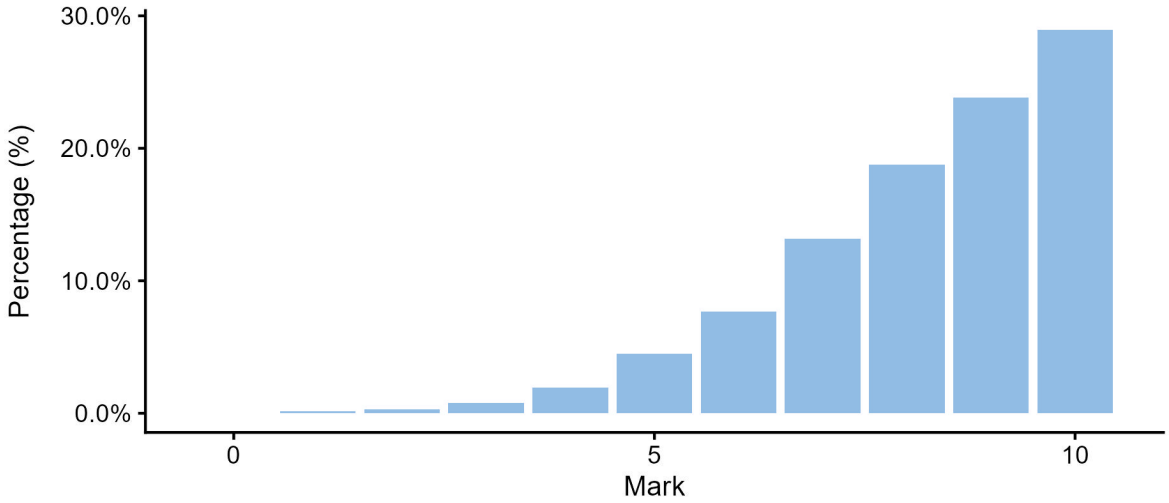
Units 3 and 4 internal assessment (IA) results

Total marks for IA

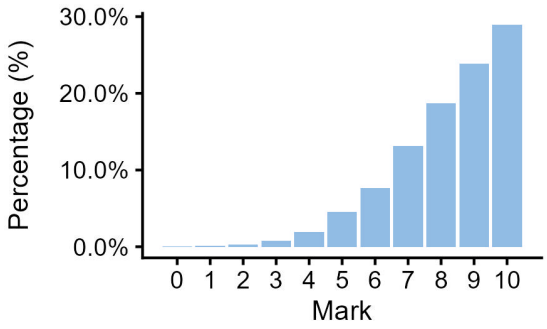


IA1 marks

IA1 total

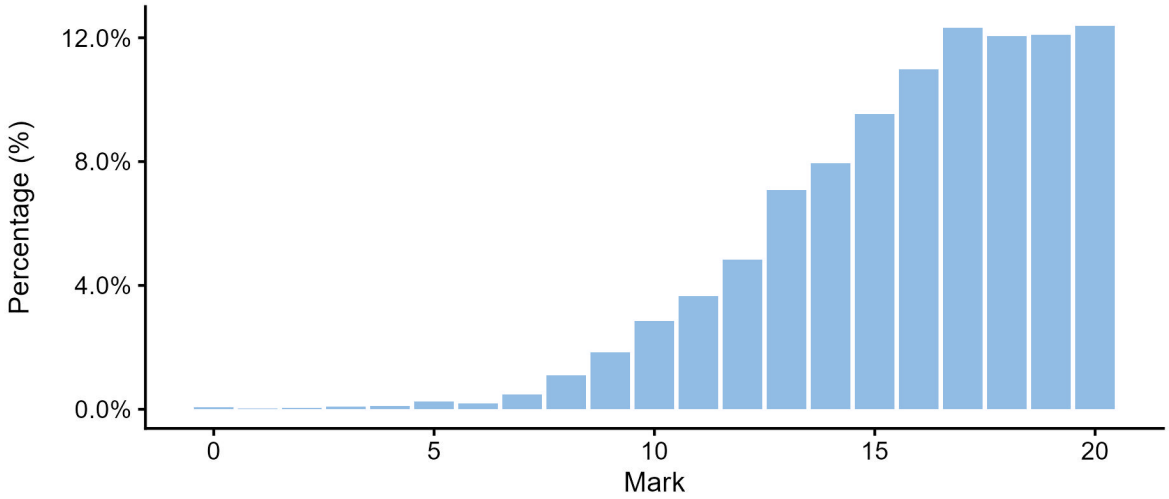


IA1 Criterion: Data test

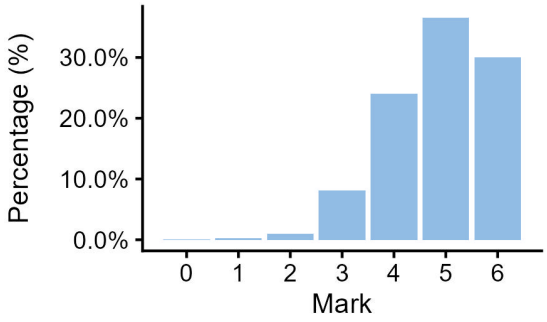


IA2 marks

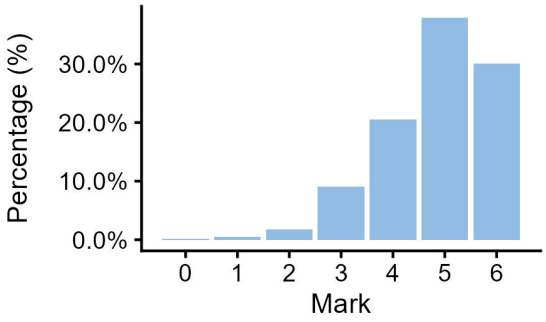
IA2 total



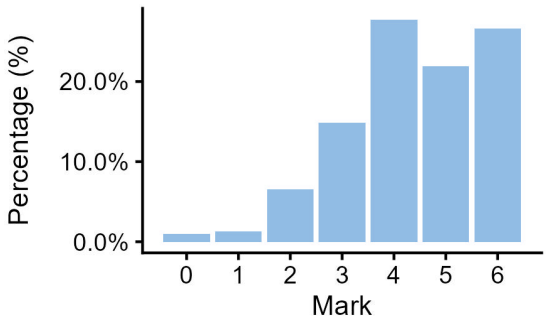
IA2 Criterion: Research and planning



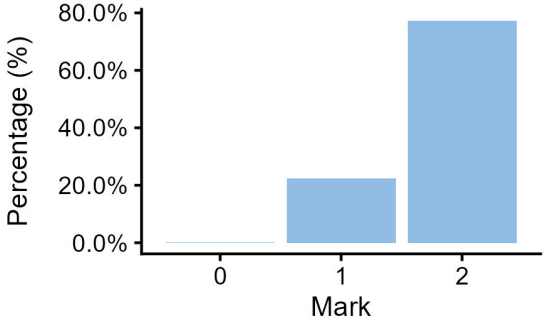
IA2 Criterion: Analysis of evidence



IA2 Criterion: Interpretation and evaluation

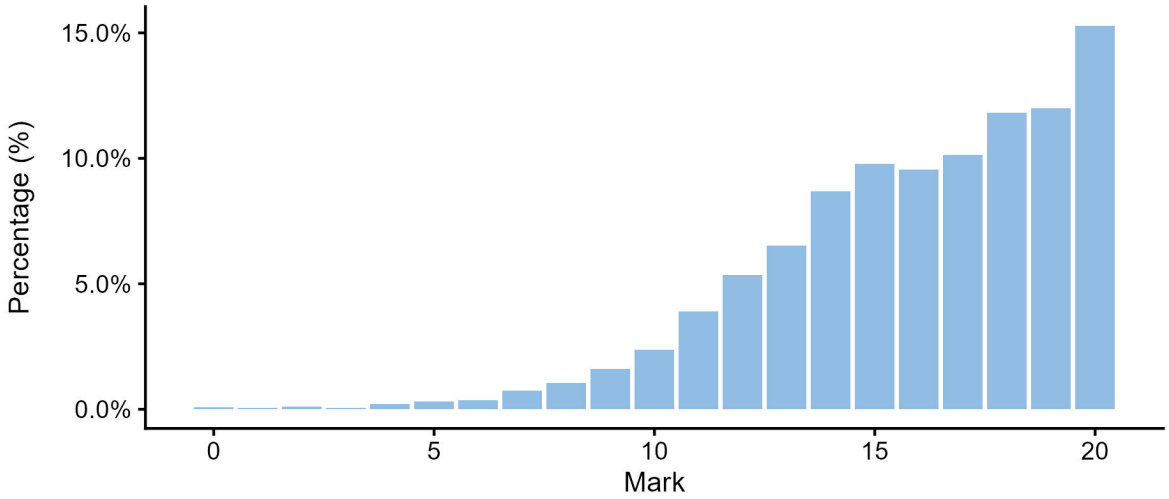


IA2 Criterion: Communication

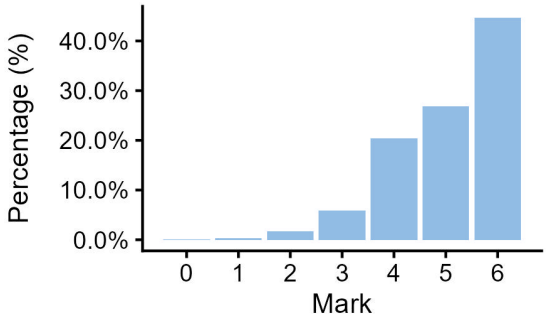


IA3 marks

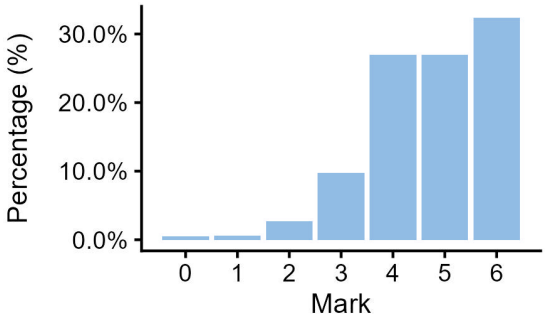
IA3 total



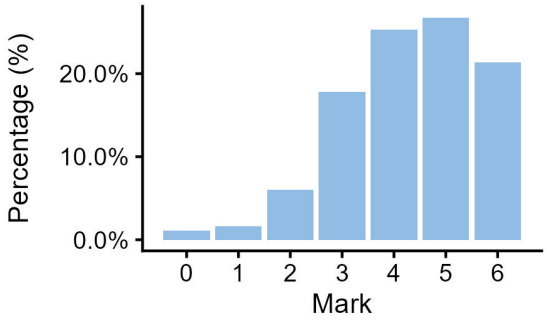
IA3 Criterion: Research and planning



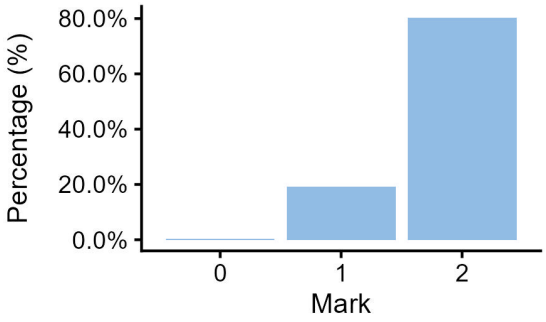
IA3 Criterion: Analysis and interpretation



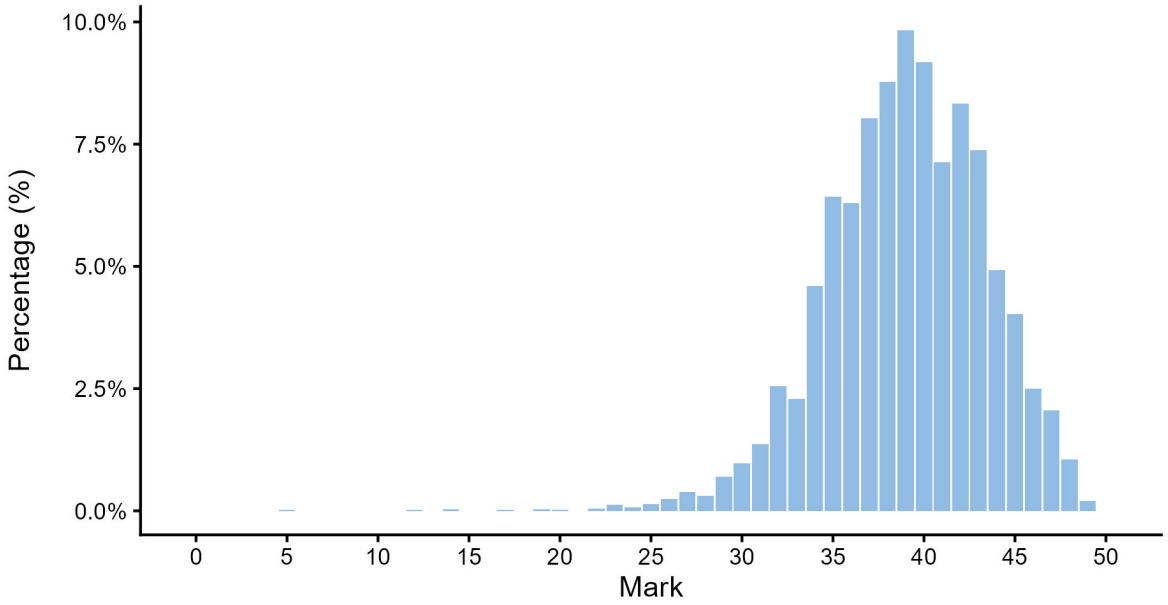
IA3 Criterion: Conclusion and evaluation



IA3 Criterion: Communication

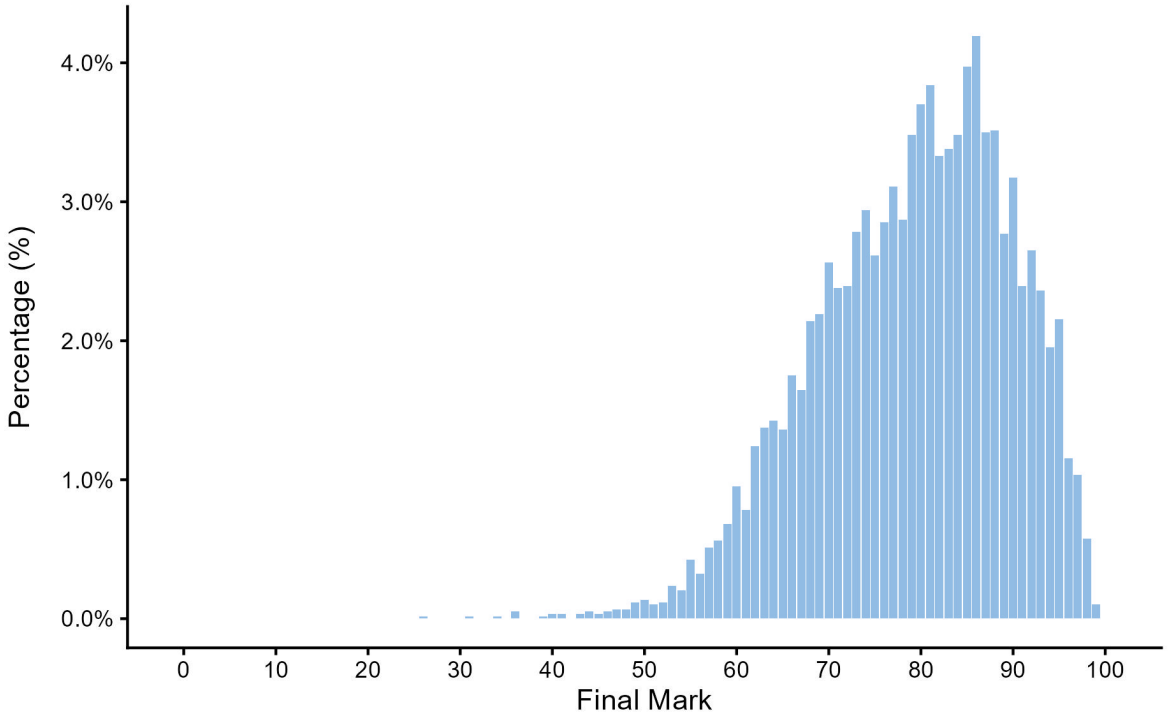


External assessment (EA) marks



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	71–50	49–20	19–0

Distribution of standards

Number of students who achieved each standard across the state.

Standard	A	B	C	D	E
Number of students	1,610	2,882	1,360	36	0
Percentage of students	27.34	48.95	23.10	0.61	0.00

Internal assessment

This 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 v7.0*, Section 9.5.

Percentage of instruments endorsed in Application 1

Internal assessment	IA1	IA2	IA3
Number of instruments	237	237	237
Percentage endorsed in Application 1	44	84	78

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 v7.0*, Section 9.6.

The following table includes the percentage agreement between the provisional marks and confirmed marks by assessment instrument. The Assessment decisions section 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	237	1,777	0	100.00
2	237	1,789	1	79.32
3	237	1,774	0	85.65

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	85
Authentication	0
Authenticity	3
Item construction	11
Scope and scale	69

Effective practices

Validity priorities were effectively demonstrated in assessment instruments that:

- included items that were aligned to Assessment objectives 2, 3 and 4 and required the use of datasets relevant to Unit 3 subject matter
- required calculation of measures of central tendency that reflected the raw data provided, e.g. the median due to obvious outliers in the data
- used cognitive verbs aligned to the focus of the question and the relevant objective, e.g. used *distinguish* to note a point of difference between two conditions in an item aligned to Objective 3.

Practices to strengthen

It is recommended that assessment instruments:

- avoid excessive repetition of data types, to allow students to demonstrate their ability to engage with a variety of datasets
- vary item types to provide students with the opportunity to demonstrate a range of skills and cognitions
- avoid asking questions about experimental methodology, which is outside the scope of the data test

- require mathematical calculations that are sufficiently complex for the syllabus conditions, e.g. adding the number of participants or subtracting two means does not reflect the scale of the task as described in syllabus conditions.

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	15
Language	13
Layout	1
Transparency	29

Effective practices

Accessibility priorities were effectively demonstrated in assessment instruments that:

- contained all necessary formulas
- labelled and organised all stimulus items logically and clearly
- were free from language errors that could affect students' interpretation of item requirements.

Practices to strengthen

It is recommended that assessment instruments:

- provide clear cues in each question to allow students the opportunity to demonstrate all aspects of the expected response, e.g. in a two-mark question that requires students to draw a conclusion and provide support for that conclusion, 'Draw a conclusion about the effect of context cues on recall. Justify your response'
- label error bars on graphs with the relevant statistic (standard error or confidence intervals, of the mean) to allow students to draw correct and relevant conclusions.

Additional advice

When developing an assessment instrument for this IA, it is essential to consider the following key differences between the 2019 and 2025 syllabuses:

- Perusal time has been modified to 5 minutes.
- The question specifications table has been revised (syllabus, p. 38). Instruments should be written in line with the revised specifications so the focus of each question aligns to the relevant objective, e.g. the cognitive verb *compare* now aligns to Assessment objective 3 as it requires identification of similarities and differences without discussion of significance.
- The revision of the question specifications table emphasises the focus of the question, rather than the verb/s used. When developing questions for a data test, using the same focus may improve alignment of questions to their objectives.

Assessment decisions

Reliability

Reliability 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	100.00	0.00	0.00	0.00

Effective practices

Reliable judgments were made using the ISMG for this IA when:

- the marking scheme clearly aligned with the endorsed assessment item
- annotations on the student script (e.g. ticks, crosses) clearly indicated where evidence for each awarded mark was found and were in alignment with the marking scheme.

Practices to strengthen

To further ensure reliable judgments are made using the ISMG for this IA, it is recommended that:

- marking schemes include expected responses that align to the cognitive verb in the question
- marking schemes clearly describe the characteristics in the response for which marks are to be awarded, alongside expected responses, e.g. provides justification for the conclusion (1 mark).

Additional advice

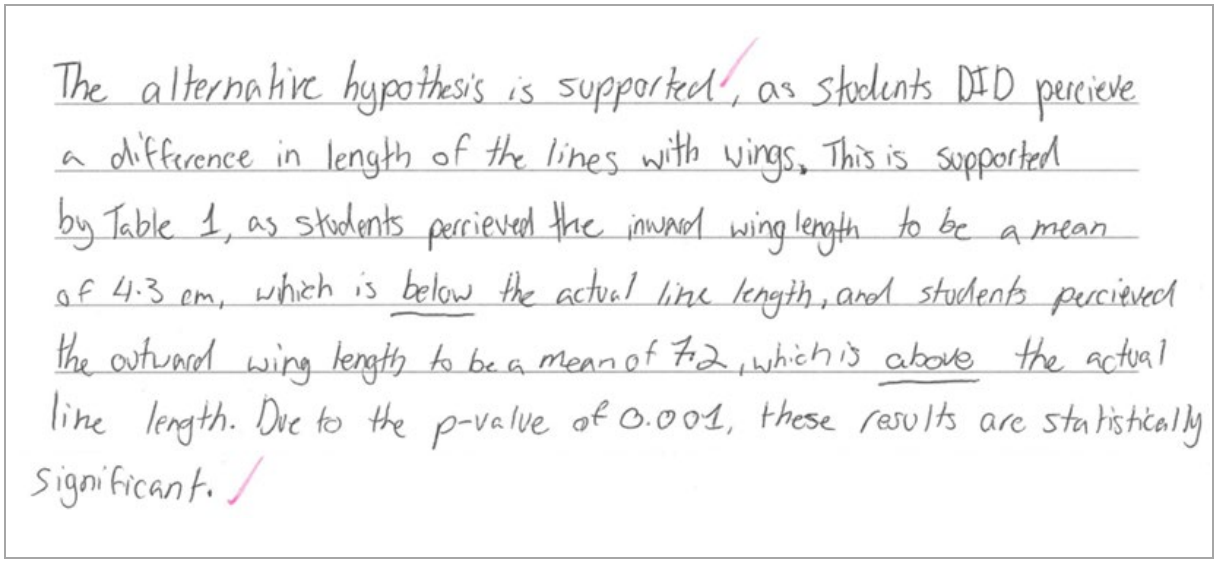
Schools should:

- ensure quality assurance practices include moderation of marking to ensure errors are corrected before provisional marks are determined
- update marking schemes to ensure alternative valid student responses are acknowledged and upload the updated marking scheme at confirmation.

Samples

The following excerpt demonstrates the effective use of annotations on a student response to indicate where evidence matches the marking scheme. The Assessment objective 4 question asks students to draw a conclusion about the findings in an experiment investigating perception. The marking scheme allocates 1 mark for drawing a correct conclusion and 1 mark for the use of appropriate evidence from the dataset to support the conclusion. The student response has been clearly annotated to show where it aligns with each of these characteristics.

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


The alternative hypothesis is supported, as students DID percieve a difference in length of the lines with wings. This is supported by Table 1, as students perrieved the inward wing length to be a mean of 4.3 cm, which is below the actual line length, and students percieved the outward wing length to be a mean of 7.2, which is above the actual line length. Due to the p-value of 0.001, these results are statistically significant.

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	24
Authentication	8
Authenticity	0
Item construction	1
Scope and scale	0

Effective practices

Validity priorities were effectively demonstrated in assessment instruments that:

- indicated which sections of the assignment would be completed as a group and which sections would be completed individually
- included experiments relevant to Unit 3 subject matter
- included specific authentication strategies, including those relevant to group work where applicable.

Practices to strengthen

It is recommended that assessment instruments:

- include all task requirements from the syllabus assessment specifications
- ensure that sample research questions provided in scaffolding cannot be used by students as a basis for their experiments
- include experiments in the task context that enable students to demonstrate systematic and effective analysis of data, e.g. experiments that
  - produce ordinal, interval or ratio measurements

- measure performance on a test of some sort, enable the calculation of descriptive and inferential statistics, the identification of trends, patterns and relationships, and the justification of conclusions.

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	2
Layout	0
Transparency	8

Effective practices

Accessibility priorities were effectively demonstrated in assessment instruments that:

- listed checkpoints in a way that clearly describes a series of stages in the development of the task
- formatted information to make it clearly legible to students.

Practices to strengthen

It is recommended that assessment instruments:

- avoid adding information that overlaps with ISMG statements about characteristics of high-level responses to the context or scaffolding of the task.

Additional advice

When developing an assessment instrument for this IA, it is essential to consider the following key differences between the 2019 and 2025 syllabuses:

- The task specifications language has been revised to align with the mid performance-level descriptor in the ISMG.

Assessment decisions

Reliability

Reliability 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	92.41	7.17	0.42	0.00
2	Analysis of evidence	86.08	12.66	1.27	0.00
3	Interpretation and evaluation	91.56	7.59	0.84	0.00
4	Communication	99.16	0.00	0.84	0.00

Effective practices

Reliable judgments were made using the ISMG for this IA when:

- for the Research and planning criterion
  - *specific and relevant* research questions and hypotheses clearly described the independent and dependent variables so the reader could understand what had been measured in data collection
  - *considered* management of ethical issues was communicated using explanations of why particular strategies were used, e.g. discussion involving informed consent for participation of minors in studies
- for the Interpretation and evaluation criterion
  - improvements and extensions to the experimental methodology were *considered* by referring directly to their effect on *related* issues with reliability and validity that were identified in the analysis of evidence.

Practices to strengthen

When making judgments for this IA for the 2025 syllabus, it is essential to consider the following key differences between the ISMGs in the 2019 and 2025 syllabuses:

- for the Analysing criterion
  - *relevant* data processing shows alignment between the *correctly* performed statistical analysis appropriate for the methodology and the graphical representations used to derive relevant trends, patterns and/or relationships in the data
  - schools familiarise students with the difference between variability and statistical uncertainty when plotting statistics as error bars.

Additional advice

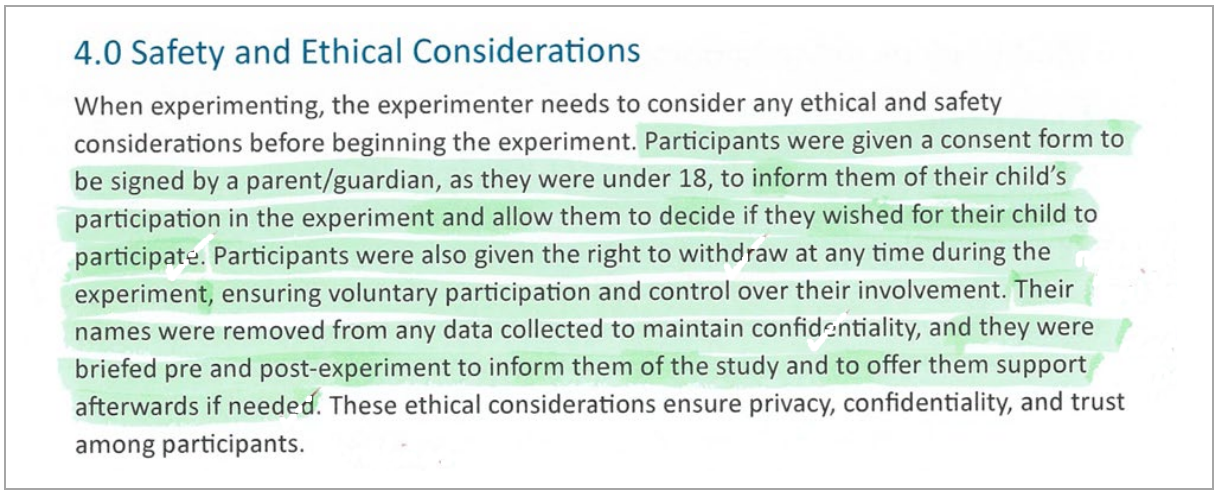
It is essential to consider the following key differences between the 2019 and 2025 syllabuses:

- The alignment between criteria and characteristics of evidence within the student response has changed; however, teachers' judgments when determining the appropriate performance level for each characteristic remain the same.

Samples

The following excerpt demonstrates *considered* management of risks and ethics by describing the strategies used and explaining the reasons why they were chosen.

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

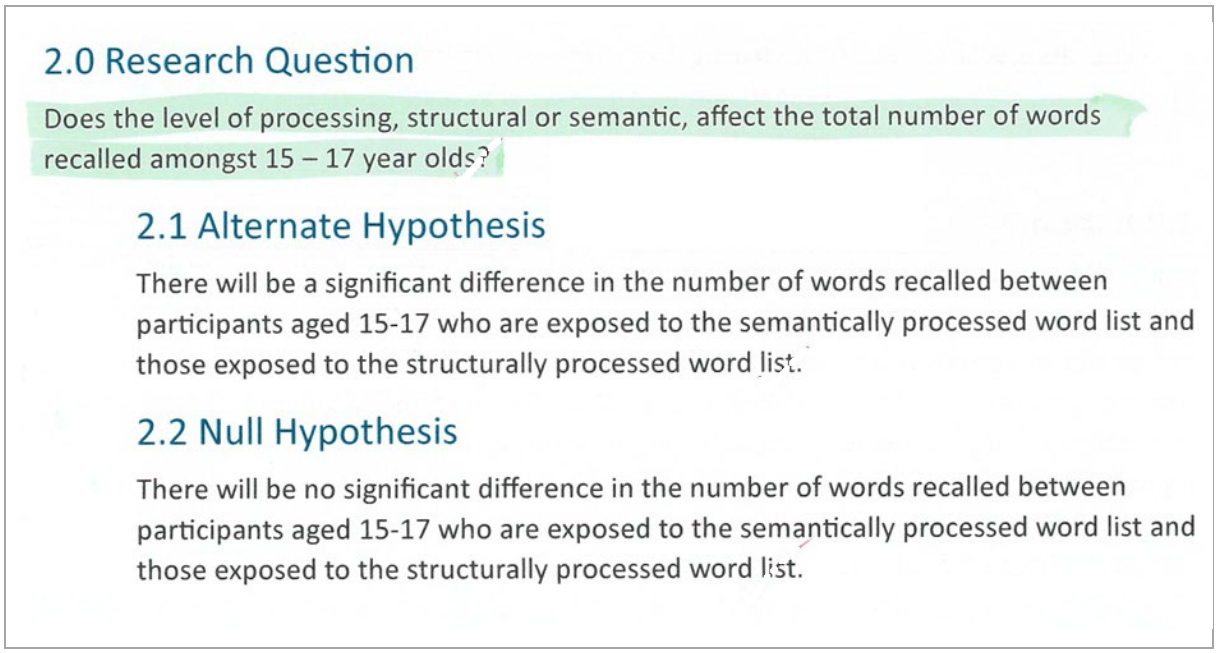


4.0 Safety and Ethical Considerations

When experimenting, the experimenter needs to consider any ethical and safety considerations before beginning the experiment. Participants were given a consent form to be signed by a parent/guardian, as they were under 18, to inform them of their child's participation in the experiment and allow them to decide if they wished for their child to participate. Participants were also given the right to withdraw at any time during the experiment, ensuring voluntary participation and control over their involvement. Their names were removed from any data collected to maintain confidentiality, and they were briefed pre and post-experiment to inform them of the study and to offer them support afterwards if needed. These ethical considerations ensure privacy, confidentiality, and trust among participants.

The following excerpt demonstrates a *specific and relevant* research question that clearly outlines the independent and dependent variables.

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



2.0 Research Question

Does the level of processing, structural or semantic, affect the total number of words recalled amongst 15 – 17 year olds?

2.1 Alternate Hypothesis

There will be a significant difference in the number of words recalled between participants aged 15-17 who are exposed to the semantically processed word list and those exposed to the structurally processed word list.

2.2 Null Hypothesis

There will be no significant difference in the number of words recalled between participants aged 15-17 who are exposed to the semantically processed word list and those exposed to the structurally processed word list.

The following excerpt demonstrates the *logical derivation* of improvements and extensions from issues identified in data analysis.

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

Figure 1

Mean Scores for Reading Silently and Reading Aloud with Standard Error Bars

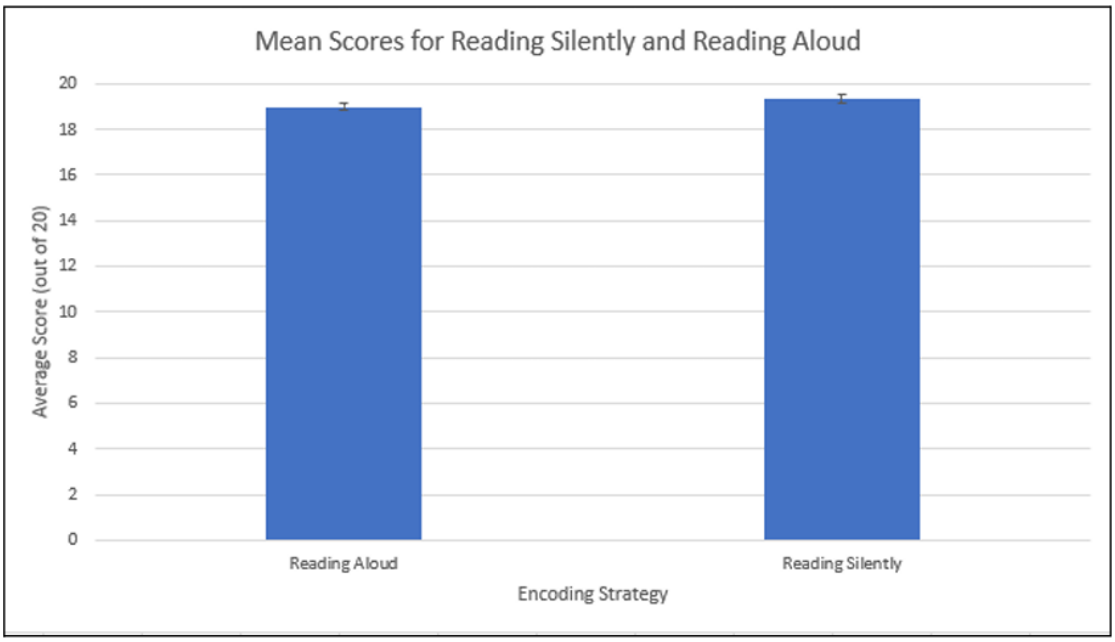


Figure 1 exemplifies how close the means were. The standard error bars above illustrate the smallness of the standard errors, suggesting that the mean accurately represents the data set. The error bars also overlap, signalling that the difference between the means is likely statistically insignificant. However, it is very likely that a ceiling effect can be seen here with the tests being too easy for participants and not sensitive enough measures. This may account for the high mean scores and the reason why 10 out of 12 participants got 20/20 (100%) on at least one of their tests.

Validity:

1. For 3 participants noticeable noise could be heard from other people in the background. This distraction may have negatively affected their results, thereby confounding the internal validity.
2. Another threat to the internal validity was the interaction of time of measurement and treatment effects. While some participants were tested in the afternoon, multiple participants were tested in the evening with P9 being clearly fatigued. As P9 had the lowest score in the R.A. condition (16/20) it is highly likely that the fatigue caused a lack of concentration.
3. A ceiling effect was also noticed among participants as a majority of the tests (13/24) scored 20/20 (100%). This indicates that the tests were not sensitive enough measures. They were too easy for participants and the results are now harder to interpret, which affects external validity, the generalisability of the study and implicates a low construct validity.

Suggested Improvements and Extensions:

Improvements:

This experiment could be refined by having:

- more participants to improve reliability by better representing the population.
- multiple tests of each participant and a classroom setting to improve test-retest reliability and ecological validity.
- an elimination of background noise and a set time for testing to improve internal validity as the confounding variables of fatigue and distraction would be eliminated.
- harder, more complex comprehension tests to eliminate the ceiling effect, providing a more sensitive measure for data collection to better discriminate between groups.

Extensions:

This experiment could be extended by having:

- a stratified sampling method with participants in age groups ranging from children to elderly. This would allow testing that could determine if 'age' is a confounding variable and how stage of life influences reading comprehension.
- modified tests that are administered based on reading ability (beginner-fluent reading level) to better understand the variable of reading ability on reading comprehension. good

These improvements and extensions could provide enhanced, more meaningful insight into the best methods of encoding for reading comprehension.

Attributions for sources quoted in excerpts

Burton, L., Saunders, J., Marangio, K., Edwards, R., Moore, V., Blaher-Lucas, E., & Ganino-Day, F. (2019). Psychology for Queensland Units 3&4. Oxford University Press.

Elias, C. S., & Perfetti, C. A. (1973). Encoding task and recognition memory: The importance of semantic encoding. Journal of Experimental Psychology, 99(2), 151–156. https://doi.org/10.1037/h0034644

Sierens, S., Gits, M., & Haryanto, S. (2018). The effects of reading strategies on comprehension performance: A review. Journal of Learning Sciences, 27(1), 83-106. https://doi.org/10.1080/10508406.2018.1536023

Topping, K. J., & Trickey, S. (2007). Collaborative learning: A review of the research. Role of Collaboration in Learning and Assessment, 25(2), 131-145. https://doi.org/10.1016/j.learninstruc.2006.08.003.

The following excerpt demonstrates *correct and relevant* processing of data using statistics appropriate for the context, that are presented as error bars to indicate either variability or statistical uncertainty.

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

Table 3: Processed results for the controlled and high impact response questions

Response	Average speed (km/hr)	Standard deviation	Standard error	Margin of error	p-value
Non-prompt	53.75	15.977	3.261	6.392	0.104
High impact	58.750	18.175	5.247	10.283	

Analysis of Evidence

Graph 1: Comparison of the Non-Prompt and Prompt Test Condition's Average Selected Speed

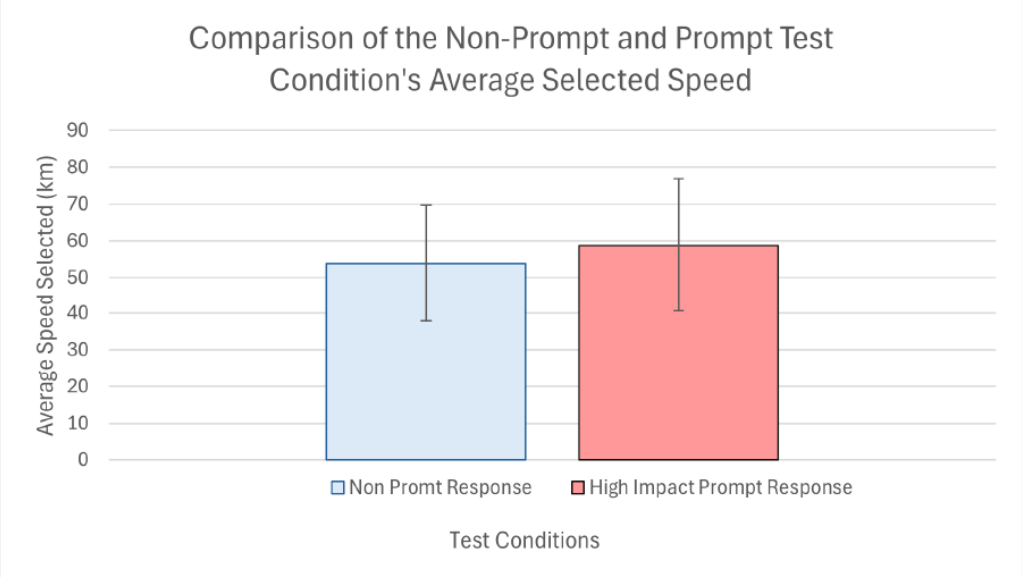


Figure 1 compares the average speed selected by participants in the non-prompt response (NPR) group and the high impact prompt response (HIPR) group, with standard deviation (SD) error bars to show the variability of the data. The non-conditioned group had an average of 53.75km/hour whereas the HIPR group had an average speed of 58.75km/hour, suggesting that the use of high modality words does increase the speed selected by participants. However, the SD error bars are elongated, indicating that the data has high variability. This conclusion is supported by the results as the SD for the NPR group is ±15.977 and the SD for the HIPR group is ±18.175, both of which are high numbers that show that the variance from the mean is great. This suggests that there were inconsistencies in the results which impacts the reliability of the experiment.

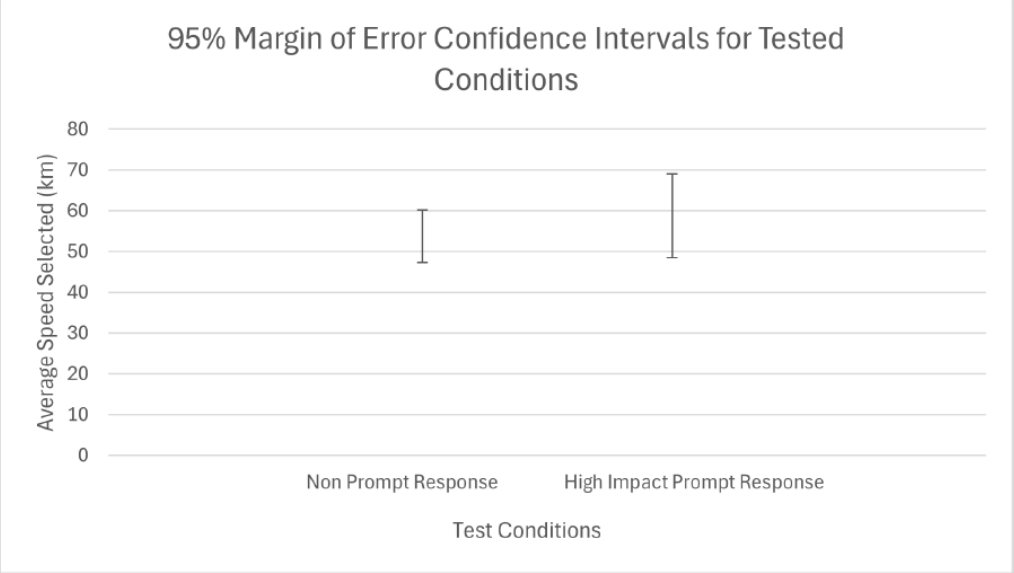


Figure 2 shows the uncertainty of the results from the tested conditions through 95% margin of error (MOE) confidence intervals. There are four outliers present within the data for the NPR group, a calculation for the upper fence of the data suggests that results past 90km/hour are outliers. There are four instances with 100km/hour being answered in the NPR as a speed by a participant, this presence of outliers indicates that there is a high variability in the results as the spread of data is not limited. To identify results without the outliers, an additional mean was calculated without the 100km/hour answers. The refined average for the NPR group is 51.739km/hour, this further assists in the suggestion that high impact words influence recalled speed as the new mean is fewer than the original NPR selected speed average.

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	9
Authenticity	2
Item construction	28
Scope and scale	4

Effective practices

Validity priorities were effectively demonstrated in assessment instruments that:

- included claims that were sufficiently broad in scope to enable students to develop unique research questions that addressed a specific aspect of the broader claim
- ensured consistency between the claims and topics listed in conditions
- used claims likely to direct students towards research questions relevant to Unit 4 subject matter.

Practices to strengthen

It is recommended that assessment instruments:

- include all task requirements from the syllabus task description
- provide claims written as single assertions with one broad independent variable and one broad dependent variable (e.g. social media is changing the nature of relationships) from which students can develop more specific research questions

- avoid claims that contain multiple assertions, e.g. the claim ‘the presence of others affects the way we think, feel and behave’ may be revised to provide a single assertion such as ‘the presence of others affects the way we behave’.

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	2
Layout	0
Transparency	3

Effective practices

Accessibility priorities were effectively demonstrated in assessment instruments that:

- used claims that were free from colloquial language and value judgments, e.g. ‘social media is changing the nature of relationships’ relates to the subject matter but avoids colloquial language and value judgments
- used claims that were free from spelling and grammatical errors.

Practices to strengthen

It is recommended that assessment instruments:

- use terms consistently between sections of the task, e.g. empirical essay or scientific essay.

Additional advice

When developing an assessment instrument for this IA, it is essential to consider the following key differences between the 2019 and 2025 syllabuses:

- The task specifications language has been revised to align with the mid performance-level descriptor in the ISMG.
- The nature of student-accessible scientifically credible sources has been clarified.
- Group elements have been added to several Forming and Finding criterion activities of the task.

Assessment decisions

Reliability

Reliability 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	95.36	4.22	0.42	0.00
2	Analysis and interpretation	95.36	3.80	0.84	0.00
3	Conclusion and evaluation	89.87	9.28	0.84	0.00
4	Communication	99.16	0.42	0.42	0.00

Effective practices

Reliable judgments were made using the ISMG for this IA when:

- for the Research and planning criterion
  - a *justified* rationale clearly showed the development of a research question by exploring the specific variables from the broader claim to be researched
  - the research question was *relevant* to Unit 4 subject matter and allowed the selection of credible sources that provided evidence to address both the research question and the broader claim
- for the Analysis and interpretation criterion
  - trends, patterns and relationships were identified through *thorough* analysis of the evidence presented across the selection of sufficient and relevant sources and were *relevant* to the research question
  - *justified* scientific arguments were developed using specific evidence from the sources.

Practices to strengthen

When making judgments for this IA for the 2025 syllabus, it is essential to consider the following key differences between the ISMGs in the 2019 and 2025 syllabuses:

- for the Forming and Finding criterion
  - research questions *specifically* address the variables used to provide a response to claim
- for the Evaluating criterion:
  - *credible* findings from the investigation are *extrapolated* to address not only the research question but also the claim
  - suggested improvements and extensions are discussed with consideration to how they are relevant to improving the investigation's ability to evaluate the claim.

Additional advice

It is essential to consider the following key differences between the 2019 and 2025 syllabuses:

- The Evaluating criterion mid-level performance in the 2025 syllabus requires students to suggest improvements and/or extensions. The Conclusion and evaluation criterion in the 2019 syllabus expected students to suggest both.
- Communication criterion characteristics from the 2019 syllabus have been split across the full range of performance levels in the Forming and Finding and the Interpreting criteria in the 2025 syllabus. Therefore, students should be aware of the requirements for communication across all performance levels.

Samples

The following excerpt demonstrates a considered rationale that clearly develops a specific research question that is relevant to Unit 4 subject matter and uses variables that are more specific than those in the claim.

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

Claim

Exposure to violent media increases aggressive behaviour.

Rationale

Media violence is any depiction of physical aggression in television, movies, music or video games (Mateski, 2023). Violence is everywhere in the media, and it is primarily in content aimed at adolescents and young adults. In psychology, aggression is behaviours that can result in both physical and psychological harm to yourself, others or property. Aggression is centred on hurting another person whether mentally or physically (Cherry, 2025). Nearly two thirds of television programs contain physical violence, Emanuel Tanay said that "Reality is distorted. If you live in a fictional world, then the fictional world becomes your reality." All the media that is consumed becomes a part of every person that watches it, he also stated that Entertainment is propaganda for violence (Kaplan, 2012). One of the most prominent forms of violent media is video games. Many video games involve gun violence, physical fights, stabbings and more. A violent video game is defined as a video game in which the options available to the player includes killing, maiming, dismembering, or sexually assaulting an image of a human being (RWJF, 2010). This is supported by the General Aggression model (GAM) which is a social-cognitive theory of aggression that proposes that violent video games increase the likeliness of a person perceiving a situation as aggressive (Burton et al., 2019). Adolescents are one of the most prominent groups playing violent video games, a study by Harvard health it was found that 66% of American teens play video games that include violence (Harvard Health, 2010). While any sort of aggressive behaviour is an issue when it is escalated to aggressive actions, thoughts and feelings the consequences become much heavier. As media violence continues to be present in various forms of entertainment, especially ones targeted at youth, understanding the effect of violent video games on adolescents' aggressive behaviours is crucial for guiding responsible media consumption. The aim of this investigation is to understand if violent video games cause an increase in aggressive behaviours and attitudes in adolescents and young adults.

Research Question

To what extent do violent video games compared to non-violent video games cause an increase in aggressive behaviour and attitudes in adolescents and young adults?

Attributions for sources quoted in excerpt

Cherry, K. (2025). Aggression Explained: What It Is and How to Recognize It. Retrieved from https://www.verywellmind.com/what-is-aggression-2794818

Gentile, D., Lynch, P., Linder, J. & Walsh, D. (2004). The effects of violent video game habits on adolescent hostility, aggressive behaviours, and school performance. Retrieved from http://www.drdouglas.org/drdpdfs/Gentile_Lynch_Linder_Walsh_2004.pdf

Kaplan, A. (2012). Violence in the Media: What Effects on Behaviour? Retrieved from https://www.psychiatrictimes.com/view/violence-media-what-effects-behavior

Mateski, M. (2023). Violence in media. Retrieved from https://study.com/academy/lesson/violence-and-the-media-how-the-mediaimpacts-violence.html

McCarthy, C. (2021). Protecting children from the dangers of virtual violence. Retrieved from https://www.health.harvard.edu/newsletter_article/violent-video-games-andyoung-people.

RWJF. (2010). Defining and Understanding Violent Video Games. Retrieved from https://www.rwjf.org/en/insights/blog/2010/10/defining-and-understandingviolent-video-games-.html
Shao, R. & Wang, Y. (2019). The Relation of Violent Video Games to Adolescent Aggression: An Examination of Moderated Mediation Effect. Retrieved from https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.0038 4/full
Willoughby, T., Adachi, P. & Good, M. (2011). A Longitudinal Study of the Association Between Violent Video Game Play and Aggression Among Adolescents. Retrieved from https://www.brockadolescentdevelopmentlab.com/uploads/1/1/2/6/112616517/w illoughby_et_al._2012_-_a_longitudinal_study_of_the_association_between_violent_video_game_play_and_aggression_among_adolescents_.pdf

The following excerpt demonstrates the justification of a conclusion, the extrapolation of credible findings to the claim, and insightful discussion of the quality of evidence.

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

Conclusion

The findings of these three studies demonstrates a clear pattern, which suggests there is a negative relationship between adherence to traditional masculine gender norms and attitude towards emotional expressivity in men. The first study conducted by Mohla and Neera produced a statistically significant p-value of 0.004 in the young-adult condition, but an insignificant relationship (0.122) between middle-aged men and emotional expressivity. The study by Cebu City supported these findings as there was a statistically significant (0.033) relationship between reduced emotional expressivity and adherence to masculine gender norms in young adult men aged 18-25. These findings suggest that younger men are more prone to emotional suppression, potentially due to peer pressure, a lessened sense of identity or a stronger desire to conform. Finally, the 2024 meta-analysis supported these findings on a larger yet more specific scale, where the results of 120 studies found that on average men are more likely to have alexithymia or difficulty describing emotions than women. This is supported by the d values of 0.22 and 0.26 respectively. This trait can potentially be attributed to environmental factors including the discouragement of

emotional expression in men. Using this evidence, the research question "To what extent does adherence to traditional masculine gender roles correlate with negative attitudes toward emotional expression in men?" can be answered. It is found that men who adhere to traditional masculine gender norms have a small but noticeable reduction in emotional expressivity, suggesting a negative attitude towards emotional vulnerability. Whilst the findings in this investigation support the acceptance of the claim 'Gender roles influence attitudes', this investigation is narrowed to only one attitude, meaning that findings cannot be accurately extrapolated to the effect of gender roles on all attitudes. Therefore, the claim cannot yet be accepted.

Evaluation

Although the third study differed from the first two due to methodology, statistical measures and the study focus, concurrent validity was achieved as all three studies produced statistically significant values that suggest a relationship between the variables. The same testing measures, self-report questionaries, were also used for each study which improves inter-rater reliability. The meta-analysis also considers the results of over 100 other studies, increasing the overall sample size and differentiation, leading to increased population validity.

However, a considerable limitation present in all three reports was the use of self-report questionnaires to collect data which decreases internal validity. This methodology potentially invites social desirability bias where participants may change their answers to align with responses deemed socially acceptable (Piedmont, 2024). This is especially present in investigations of sensitive topics including mental health, where participants may be less inclined to disclose the truth. To improve this limitation, quantifiable measures should be used as an alternative to qualitative and subjective rankings on a Likert scale.

The next limitation occurring across the first two studies relates to population validity. These investigations utilised purposive and convenience sampling and relatively small sample sizes. This creates potential for cultural and environmental bias which reduced the reliability of results and their ability to be generalised. To improve this limitation, random sampling should be used to eliminate participant bias.

Finally, the meta-analysis differs from the first two studies due to the methodology being the combination of results from over 100 studies. This creates limitations due to secondary data, however it also reduces population validity concerns, and increases concurrent and inter-rater reliability. This analysis also focuses on the specific traits of alexithymia which differs from generalised reduced emotional expression in the first two studies, however results support the findings of the other investigations. Different testing measures were also used for many of the primary investigations, potentially reducing construct validity. Finally, different statistical measures were used across the three studies making direct comparison more difficult.

Attributions for sources quoted in excerpts

Mohla, D., & Neera, D. (2023). Men Are Stoic, or Are They?: Emotional Suppression in Men and its Relation with Masculinity Norms. The International Journal of Indian Psychology. https://ijip.in/wp-content/uploads/2023/09/18.01.399.20231103.pdf

Mahinay, C., Miano, E., Sarausad, E., & Abejar, M. (2024). Conformity of Masculine Norms as a Predictor of Negative Emotional Expression Among the Cisgender Male College Students of Cebu City. Chelonian Conservation and Biology. https://www.acgpublishing.com/index.php/CCB/article/view/940/1067

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. The external assessment papers and the EAMG are published in the year after they are administered.

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 (29 marks)
- Paper 2, Section 1 consisted of short response questions (46 marks)

Assessment decisions

Assessment decisions are made by markers by matching student responses to the external assessment marking guide (EAMG).

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	10.25	**50.70**	13.79	24.77
2	11.39	**82.89**	0.43	5.13
3	8.31	5.57	**56.53**	29.32
4	0.85	2.40	**94.55**	2.04
5	0.41	2.77	22.61	**73.95**
6	**49.39**	17.37	2.75	30.29
7	11.93	1.97	5.70	**80.17**
8	**51.63**	10.79	20.40	16.87
9	8.57	28.83	**36.66**	25.74
10	6.55	**71.10**	13.85	8.16
11	12.43	**65.81**	6.91	14.57

Question	A	B	C	D
12	**44.06**	24.30	18.70	12.60
13	**77.82**	6.66	10.04	5.25
14	8.92	9.04	6.19	**75.64**
15	9.28	34.52	6.93	**48.89**
16	27.33	0.92	**65.15**	6.36
17	**83.61**	12.10	0.54	3.56
18	17.60	12.77	**65.72**	3.63
19	4.63	5.70	4.62	**84.79**
20	**68.66**	3.82	8.12	19.10

Effective practices

Overall, students responded well to:

- items requiring
  - identification of data from tables and graphs
  - interpretation of tabulated or graphical data to identify trends, patterns and relationships, particularly with reference to correlational data
  - a description of key concepts in simple contexts.

Practices to strengthen

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

- provide opportunities for students to
  - explain concepts outlined in the subject matter and provide specific relevant examples
  - analyse and interpret evidence to identify trends, patterns and relationships in data and provide appropriate justification for conclusions from presented data
  - respond to cognitive verbs, such as *deduce*, *infer*, *contrast* and *justify*
  - use the number of marks allocated to a question to indicate the features required in the response.

Samples

Short response

Question 26b) from Paper 1

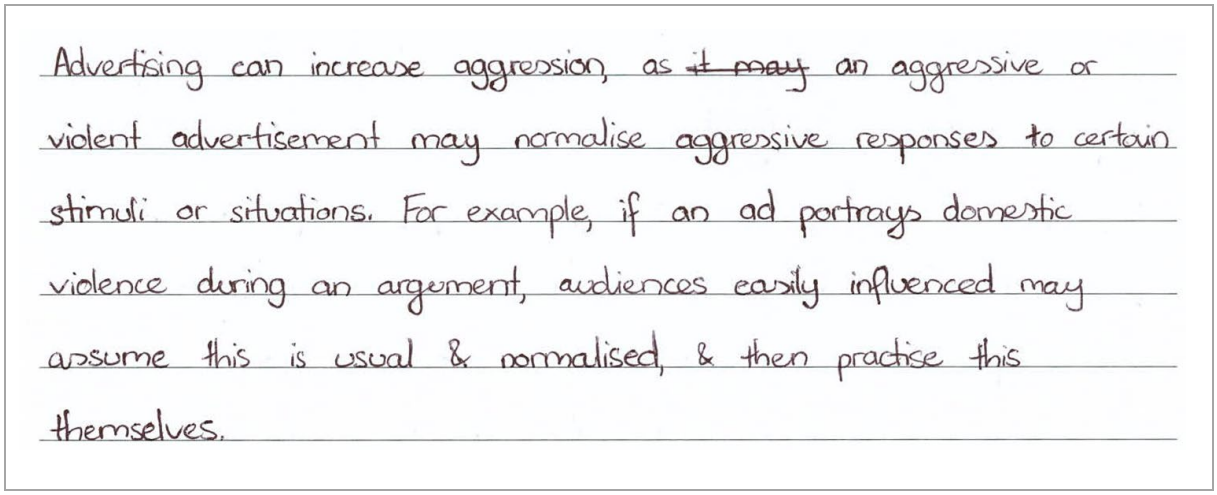
This question required students to explain how advertising could affect aggression and provide an example for 2 marks.

Effective student responses:

- offered detail about the way that advertising could affect aggression, beyond a simple description of the phenomenon
- provided a specific example of the phenomenon.

This excerpt has been included:

- because it explains the effect by including details about the mechanism of causation
- it provides a specific example.


Advertising can increase aggression, as ~~it may~~ an aggressive or violent advertisement may normalise aggressive responses to certain stimuli or situations. For example, if an ad portrays domestic violence during an argument, audiences easily influenced may assume this is usual & normalised, & then practise this themselves.

Question 27 from Paper 1

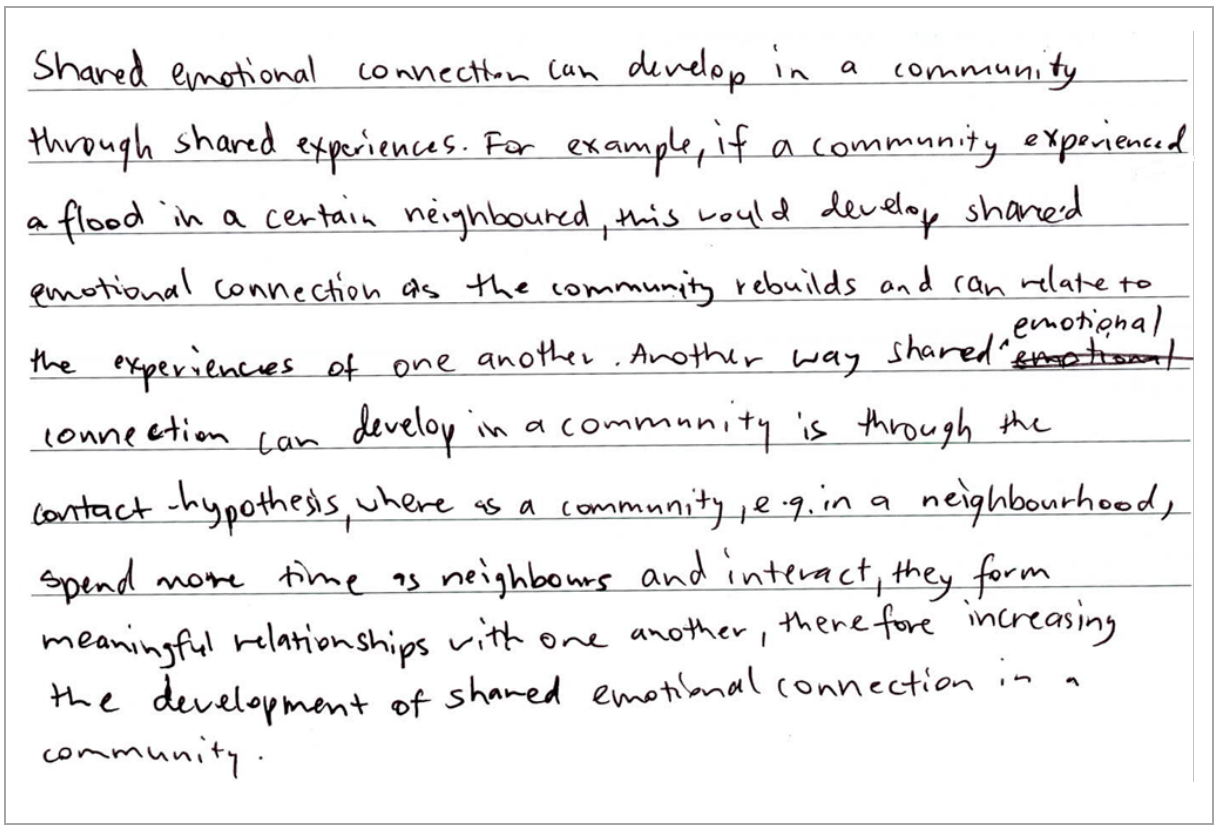
This question required students to describe two ways a shared emotional connection can develop in a community, in accordance with McMillan and Chavis (1986), for 2 marks.

Effective student responses:

- described two distinct ways that shared emotional connection can develop in a community
- were aligned to the McMillan and Chavis (1986) study.

This excerpt has been included:

- because it accurately describes two ways that a shared emotional connection can develop in a community, according to the McMillan and Chavis (1986) study.


Shared emotional connection can develop in a community through shared experiences. For example, if a community experienced a flood in a certain neighboured, this would develop shared emotional connection as the community rebuilds and can relate to the experiences of one another. Another way shared emotional ~~emotional~~ connection can develop in a community is through the contact hypothesis, where as a community, e.g. in a neighbourhood, spend more time as neighbours and interact, they form meaningful relationships with one another, therefore increasing the development of shared emotional connection in a community.

Question 1c) from Paper 2

This question required students to infer the results of an experiment if bystander apathy was the only variable affecting behaviour, and to contrast these results with those found in Darley and Latane's (1968) experiment investigating the impact of bystander numbers on helping behaviours. The question was worth 2 marks.

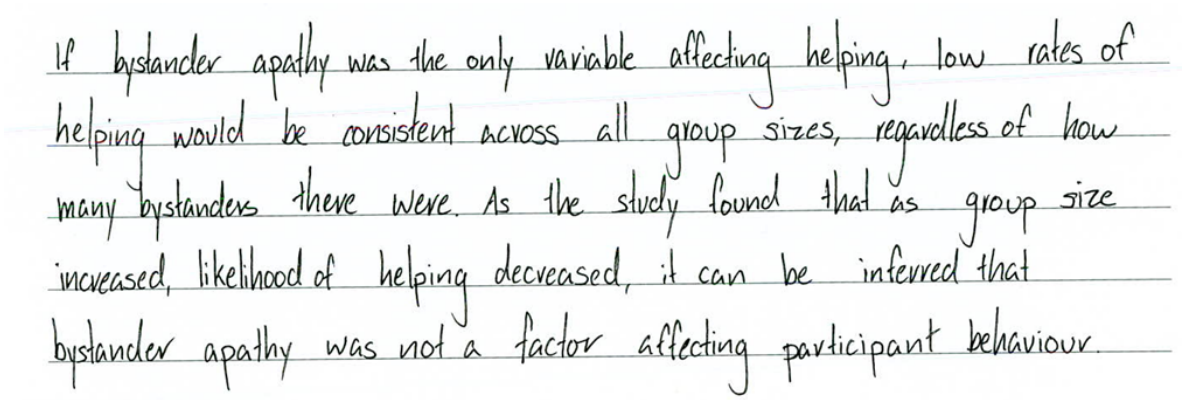
Effective student responses:

- inferred appropriate experimental results that would be consistent with indifference
- clearly contrasted the inferred result with the actual result.

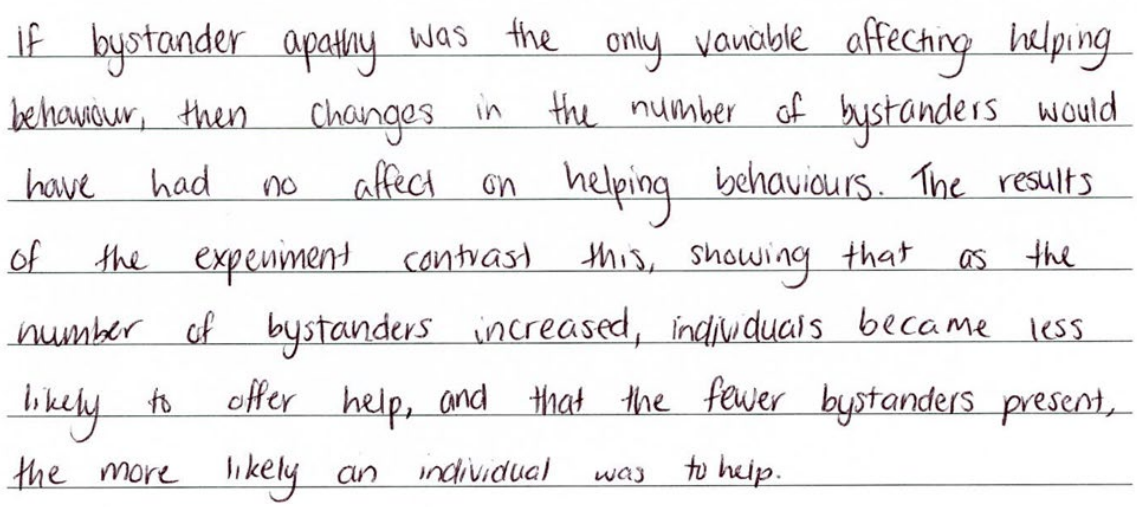
These excerpts have been included:

- to demonstrate appropriate inferences
- to demonstrate explicit contrasts between inferred and actual results.

Excerpt 1



Excerpt 2



Paper 2, Question 5c

This question required students to draw a conclusion from presented data about the effect of context-dependent cues on retrieval of meaningful information and justify the response. The question was allocated 2 marks.

Effective student responses:

- drew a suitable conclusion from the presented data
- justified the conclusion appropriately.

This excerpt has been included:

- to demonstrate a suitable conclusion that aligned to the presented data
- to demonstrate appropriate justification using the result of an inferential test of significance.

It can be determined that the precense of context-dependent cues increases the likely hood of successful recall. Those tested in silent conditions that matched with the silent studying conditions produced a higher mean result than those who tested in mismatched conditions. As the p-value for this experiment was below the assumed set alpha of 0.05 it can be determined that these results are statistically significant.