Internal assessment

Component grade boundaries

Grade:	1	2	3	4	5	6	7
Mark range:	0 - 3	4 - 6	7 - 10	11 - 13	14 - 16	17 - 19	20 - 24

The range and suitability of the work submitted

Many schools presented a very large range of inventive and original investigations. These were a real pleasure to read. Nevertheless, examiners reported that there appeared to be a greater reliance on classic investigations, some of which are prescribed in the core of the program, with little or no attempt to modify them.

Overall, work was of a suitable standard.

Consideration of safety and ethics were frequently lacking particularly in work with microbes.

There were some trivial investigations that were not of the appropriate level for the IB biology course.

Candidate performance against each criterion

The application of the assessment criteria by teachers was generally good, though often overgenerous, sometimes very generous. There are cases where teachers pointed out significant weaknesses in a criterion but then chose to award the highest grade. Therefore, more rigor is necessary when applying the final mark. Teachers were rarely considered too severe.

Evaluation is still the weakest criterion for many candidates. This criterion is difficult and it does discriminate between the candidates. For many candidates, Analysis was also a criterion that needed more attention. Many were happy to leave the processing at the level of calculating means only.

Personal engagement (PE)

Some form of personal significance was expressed in most cases. While many were clearly inspired by an observation or an issue, many were contrived (for example, "I have always been interested in..."), or there was no expression of personal significance at all.

The originality of the exploration was mostly acceptable, sometimes exceptional. There were, however, too many cases of classic investigations being used with little or no attempt to modify them.

Personal input is evident in persistence to collect data, in research into the background, when establishing the scientific context of the conclusion, in following through the investigation and in the choice of methods of analysis. Once again, this was clearly evidenced by many candidates. For others, it seemed that after a good start with an interesting research question, they failed to follow through.

Personal input can be reflected at the simplest level by having completed the investigation, but those following classic experiments, with no sign of application, cannot expect to score highly. There must be some indication that there is a commitment to the investigation.

A number of examiners observed that teachers seemed to be content awarding 2 marks for statement of purpose at the beginning of the report. Teachers need to look further for evidence before making a judgment on this criterion.

When assessing this criterion, teachers should look out for the following:

- A statement of purpose;
- The relationship with the real world;
- The originality of the design of the method (choice of materials and methods);
- Evidence of trial runs;
- The difficulty of collecting data (evidence of tenacity);
- The quality of the observations made;
- The care in the selection of techniques to process the data;
- The reflections on the quality of the data;
- The type of material referred to in the background or in the discussion of the results;
- The depth of understanding of the limitations in the investigation;
- The reflections on the improvement and extension of the investigation.

Marking this criterion requires a holistic approach and will almost certainly overlap with components of other criteria.

Exploration (EX)

For many submissions, the research question lacked sufficient focus. Scientific names were not always used and the range of the independent variable was not always given. For example, a candidate whose question read, "How will different amounts of sugar have an effect on cell respiration in yeast used in bread making?" should have considered including the species of yeast and the sugar used. The word "amount" could have been made more specific by substituting with "mass", or "volume" or "moles". The range of sucrose concentrations to be used should have been indicated. A research question can also include how the measurements will be taken by introducing the dependent variable.

The requirements for the background are that it needs to be focused and contain relevant information that is clearly linked to the research question. There were many cases of superficial or irrelevant material. The independent variable needs to be justified. The dependent variable needs to be explained. The discussion of controlled variables is needed to demonstrate that the candidate appreciates the other factors that may have an impact on the experiment. Uncontrolled variables, for example, room temperature, which may have a significant impact on the experiment need monitoring. One cannot just assume that carrying out the experiments in the same place is enough. Control experiments needed to be considered more frequently.

The methods were either written in prose or recipe-style. Both were acceptable. Where the method was not clear, this has an impact on both Exploration and Communication criteria. The weaker submissions were seen mostly from candidates who investigated a topic in which causal relationships were difficult to confirm and a large number of controls were missing. For example, human physiology studies, with limited data sets and poorly controlled variables.

Examiners found the candidates' understanding of concentration to be relatively weak. A serial dilution of a molar solution of sucrose would sometimes result in the stock solution being identified as 100% sucrose.

When assessing this criterion, teachers should look out for the following:

- The protocol for collecting the data;
- The range and intervals of the independent variable;
- The selection of measuring instruments (where relevant);
- Techniques to ensure adequate control (fair testing);
- The use of control experiments;
- The quantity of data collected, given the nature of the system investigated;
- The type of data collected;
- Provision for qualitative observations.

Safety, ethics and environmental impact needed to be addressed, or further explained, in a large number of investigations. It is true that some investigations do not have any issues in these areas but there were plenty that did, yet candidates showed little or no evidence of concern. It is not sufficient to identify potential areas where safety is an issue, an indication of how the issue is to be avoided should also be seen.

There were some microbiological methods carried out by candidates that were very inappropriate for a school environment. Feedback has been provided to the schools who did not address this.

There were some potentially dangerous practices in investigations, for example, the use of animal manure samples.

The use of consent forms with human volunteers is not being universally adhered to. This is an essential ethical practice.

The environmental impact and safety for fieldwork was often ignored or treated superficially.

The following guidelines should be applied: Labs using bacteria

- Only culture known, non-pathogenic strains of microbes. For example, do not culture from hands or swabs of door handles.
- Bacterial cultures should be obtained from a reputable supplier or from a safe source such as a university microbiology laboratory.
- Do not test for antibiotic resistance. There are enough antibiotic resistant strains circulating in the environment without more being selected.
- Apply strict rules of hygiene and aseptic techniques.
- Do not culture microbes at 37°C. Incubation should be carried out below 30°C.
- Always label plates so they can be clearly identified; never open them for inspection.
- Do not tape all the way round a Petri dish. This encourages anaerobic conditions.
- Never assume that what is growing in the culture is the strain that was inoculated, even if non-pathogenic strains have been used.
- Always sterilise used cultures and dispose of the cultures following local health and safety regulations.

When assessing safety, ethics and environmental issues, teachers should look for the following:

- Evidence of a risk assessment;
- An appreciation of the safe handling of chemicals or equipment (e.g. the use of protective clothing and eye protection);
- Consideration of basic hygiene;
- The application of the IB animal experimentation policy;
- A reasonable consumption of materials;
- The use of consent forms in human physiology experimentation;
- The correct disposal of waste;
- Attempts to minimise the impact of the investigation on field sites.

Analysis (A)

The presentation of raw data was generally accurate, however, qualitative observations were missing from many submissions. Qualitative observations are expected to accompany the raw data. Their impact will depend upon the nature of the investigation, for example, fieldwork should always have a site description which could take the form of maps, sketches or photographs with annotations.

Raw data from data logging may be expressed as a graphical readout. It should be accompanied by the necessary information, such as units and degrees of precision (if relevant) in the axis titles. A candidate should only present a representative sample of the raw data, for example, when large amounts of data have been collected using data logging. A representative graphical readout revealing how data is derived is acceptable. In this way, the derived data becomes the raw data.

The processing of data varied between schools. Most candidates managed the basics, for example, means and standard deviations. However, there were still candidates who tried to apply a standard deviation to a sample size that was too small ($n<5$).

There were cases where candidates calculated mean rates by averaging the data for all the trail runs combined, then calculated the mean following this. This is inexact. The rate for each run needs to calculated individually, followed by the mean from all rates.

Candidates are still confusing R^2 with the correlation coefficient r. R^2 is the coefficient of determination. R^2 can be used as an indicator of the goodness of fit of a trend line. It can approximate to the product moment correlation coefficient (r) if the trend line is straight but it is always a positive value, unlike the correlation coefficient which can be negative.

- Several candidates used significance tests from t-test to ANOVA. Although good, they need to be appropriately applied and there needs to be sufficient explanation for the processing to be followed. The use of programmes, such as Microsoft Excel, which produce a statistic, such as a p-value or a correlation coefficient, is fine, however, the candidate needs to know what the value actually represents.

  >30 is considered a large sample;

  15-30 a small sample;

  5-15 a very small sample;

  <5 is usually considered too small a sample to apply tests like the t-test.

- Rates and proportions were not always calculated where they were appropriate.

- In some cases, measurement uncertainties were presented but not discussed. Candidates are expected to appreciate the limitations of their instruments and, where they have a choice, to select the most appropriate one. In biology, the biggest issue for uncertainties is in the variation in the biological material (expressed as standard deviations, standard error or max-min range). Error bars showing variation were frequently used on graphs but their significance, or even what they represented, was often absent. In other cases, the error bars were incorrectly placed or they had no bearing on what the candidate had calculated.

The interpretation of the data was often well presented after each set of data. However, it was sometimes mixed with the conclusion. The use of statistics may have been satisfactory but they were not always interpreted well. As with calculators, the use of a program like Excel is useful but can lead to accepting values without truly understanding them. Huge mistakes can result from this (for example, confusing the t-statistic with the p-value), leading to an erroneous conclusion. Often, interpretation was handicapped by the limited degree of data processing.

Evaluation (EV)

This was the weakest criterion for many candidates. Although is it a difficult skill for some candidates, it seemed that these were rushed in an attempt to finish off the report. Schools may want to consider further the impact of the conflicting deadlines between each chosen subject, theory of knowledge and extended essays and other requirements of the candidate.

Conclusions were not always supported by the data and some explanations were missing. Some candidates were rather overoptimistic in their conclusions. Clearly the data did not fully support the conclusion made but they would aim to put a positive spin on it. A scientific context is needed for a full discussion and this was frequently superficial or absent. For weaker candidates, the conclusion was just a description of the results. Many examiners commented that candidates were correctly interpreting statistical significance tests but they were not referring back to the research question.

- The evaluation of methodology was still a challenge to most candidates. The consideration of the strengths was frequently missed. Weaknesses were often restricted to practical details or sloppy manipulation and the level of impact on the conclusion was often not discussed. Proposed improvements were sometimes unrealistic and often too vague. Extensions were often missed or illogical, not following on from the investigation. This was an area where examiners felt that teachers were often marking over-generously.

When assessing this criterion, teachers should look for the following:

- A discussion of the strengths – this might be quite general or implicit or it might refer to specific parts that worked well or data that was consistent;
- Discussion of the reliability or the data;
- Identified weaknesses in the method and materials;
- The evaluation of the relative impact of a weakness on the conclusion.

Communication (C)

The responses to the communication criterion were generally good. Those who communicated well were candidates who had already scored highly in the other criteria.

The most common problems in the work were:

- The use of whole pages for titles – this is not necessary.
- Whole pages for a list of contents – again, this is not necessary.
- Blank data tables presented at the end of the method section.
- Repetitive tables, when one representative sample would be sufficient.
- Tables split over two pages, or with a title on one page and the table or graph on the next.
- Multiple graphs drawn when these could have been combined. This not only saves space but it also improves and may aid candidates to make better comparisons.
- Squashed graphs so the distribution of the data was difficult to judge. This was often due to the candidates failing to reformat the selected font.
- Bibliography, footnotes, endnotes or in-text citation missing.
- References with an incomplete format. Often just a URL is given.
- Inefficient data table headers. The art of designing data tables needs to be taught. A hand drawn sketch of the table layout should be considered first.
- Scientific nomenclature was not always used and the formats were not always respected.

For graphs that result from data logging that are used to derive a value, for example, a rate, one example can be presented to explain the processing then the rates derived can be organised in a table.

The format for the citations, when they were presented, was mostly correct.

Format of scientific names was sometimes incorrect (small case letter for species name and the name should be presented in italics).

Units were occasionally missing and use of non-metric units, for example, teaspoons and cups, were noted by some examiners.

Measurement uncertainties were occasionally missing.

The numbers of decimal places were sometimes irregular or did not correspond to the precision of the data.

In general, the reports were of suitable length.

No automatic penalties were issued for reports that were slightly longer, as long as reports remained relevant and concise, as detailed in the Communication criteria.