Full Evaluation of the IB Mathematics AA SL Exploration

Formulating a Suitable Equation to Model the Cooling Process of Hibiscus Tea

Strict teacher review under the current assessment model
Review date: 27 August 2026
File reviewed: 22 PDF pages, including a two-page bibliography
Course: IB Mathematics: Analysis and Approaches, Standard Level

1. How the final IB Mathematics AA SL grade is calculated

For the current Mathematics AA SL course, the final subject grade is assembled from three weighted components. The exploration is important, but it is only one fifth of the subject result.

Component	Conditions	Weight in the final subject result
Paper 1	1 hour 30 minutes; no technology allowed	40%
Paper 2	1 hour 30 minutes; technology allowed	40%
Internal assessment: mathematical exploration	Assessed by the teacher and externally moderated	20%
Total		**100%**

If P_1, P_2 and I denote percentage scores for Paper 1, Paper 2 and the exploration, the weighted subject total is

$$S = 0.40\,P_1 + 0.40\,P_2 + 0.20\,I\,.$$

For example, Paper 1 at 80%, Paper 2 at 80% and an exploration at 15/20, or 75%, would give

$$S = 0.40(80) + 0.40(80) + 0.20(75) = 79\,.$$

This is only an illustration. The final grade from 1 to 7 is awarded from the combined session result using the grade boundaries for that examination session. Therefore, neither 15/20 nor 18/20 on the exploration guarantees a 7 by itself.

How the exploration's 20 raw marks are formed

The exploration is marked out of 20. Its five current criteria are:

Criterion	Maximum	What it rewards
A: Presentation	4	A coherent, well-organised and concise exploration
B: Mathematical communication	4	Relevant, appropriate and consistently correct mathematical language, notation and representation
C: Personal engagement	3	Authentic, independent engagement that drives the investigation
D: Reflection	3	Substantial critical reflection throughout the exploration
E: Use of mathematics — SL	6	Relevant and correct mathematics at an appropriate level, supported by thorough understanding

Criterion	Maximum	What it rewards
Total	**20**	The raw exploration mark, scaled to 20% of the subject result

The current five-criterion structure remains the correct basis for work submitted before the new course's first assessment in May 2029. The redesigned criteria—Problem specification, Abstraction, Computation and Interpretation—belong to the course first taught from August 2027 and first assessed in May 2029; they should not replace the present criteria for a current exploration.

2. Overall judgement

This exploration begins from a worthwhile idea. It has a credible personal context, original experimental data and a sensible connection between exponential functions, logarithms, regression, differential equations and model criticism. That is a stronger foundation than many average explorations.

In its present form, however, several important mathematical inconsistencies prevent a secure top mark:

1. Equation C is printed once with $e^{-0.57955t}$ instead of $e^{-0.057955t}$, a factor-of-ten error.
2. The parameters of equation B do not agree across the regression graph, prose, displayed equation, prediction table and RMSE calculation.
3. The alternatives are not evaluated on the same observations. The conclusion that C is the best equation for the relevant domain is contradicted by a fair calculation from the displayed data.
4. The adjusted RMSE of $1.991\,\square^{\circ}C$ cannot be reproduced. Table 1.7 gives approximately $2.050\,\square^{\circ}C$ for $t \geq 10$.
5. Boyle's law is used incorrectly to explain the approach to room temperature.
6. Units, symbols, axes and numerical precision are inconsistent in several places.

My strict but realistic estimate is **15/20**, with a plausible moderation range of **14–16/20**. After a mathematically coherent revision, **18–20/20** is attainable; **19/20** is a realistic high target. The exploration itself does not receive a standalone grade from 1 to 7.

3. Criterion-by-criterion estimate

Criterion	Current estimate	Evidence for the estimate	What is needed for the top band
A: Presentation	3/4	The progression from question to conclusion is generally clear and most figures are placed near the relevant discussion. It is not fully concise: three very similar parameter estimates occupy substantial space, numbering is inconsistent, the page count is wrong and navigation is weak.	Remove repetition, combine raw-data tables, add page numbers, standardise references, and make the model-selection logic shorter and more direct.
B: Mathematical communication	2/4	There is a good range of representations, but serious inconsistencies remain: $t=22.4$ instead of $T=22.4$, seconds instead of minutes, the factor-of-ten exponent error, missing parameter units, two meanings for m, and conflicting regression parameters.	Use one notation table and one reproducible source for every parameter. Check every equation, axis, table, label and unit.

Criterion	Current estimate	Evidence for the estimate	What is needed for the top band
C: Personal engagement	2/3	The tea routine, chosen threshold and self-collected measurements are authentic. Some development reads as reconstructed for the assessment, particularly the staged sequence from A to B to C and the selection of two favourable points after viewing the data.	Show a genuine mathematical decision: state a prediction, use residuals to reject or refine an estimate, and reserve at least one trial for validation.
D: Reflection	3/3, narrowly	There are recurring reflective points about transformation, an impossible intercept, limitations and model comparison. The strength is reduced because an acknowledged unfair comparison is still used to justify the final choice.	Act on each criticism mathematically and quantify its consequence instead of listing limitations without testing them.
E: Use of mathematics — SL	5/6	The mathematics is relevant and mostly correct. The differential-equation derivation is above the minimum expected at SL and is largely understood. The wrong exponent, inconsistent regression, incorrect RMSE and unsupported model choice prevent the work from being consistently correct.	Add a direct nonlinear fit, compare methods fairly, show reproducible residual and RMSE calculations, include a sensitivity or uncertainty analysis, and remove all notation errors.
Total	**15/20**	A strong investigation with material errors that affect the central conclusion.	**Realistic revised target: 19/20.**

4. Numbered error and revision table

4.1 Form, structure, data collection and presentation

No.	Location	Error or risk, and why it matters	Exact correction or possible replacement wording
1	Title page	“15 pages” does not match the submitted file. The PDF has 22 pages; excluding the two-page bibliography leaves 20. An incorrect count suggests that the final document was not checked and weakens presentation.	Recount after the final edit. For example: **“20 pages excluding the bibliography.”**
2	Whole document	Visible page numbers are missing. This makes a long exploration harder to navigate, moderate and reference.	Add page numbers in the footer. The title page may hide its number, but should normally still be counted.
3	Title page	“An exploration into Cooling Process” is grammatically incomplete and unnatural.	Use **“An exploration of the cooling process of hibiscus tea”**, or remove the line.
4	Whole document	At 20 pages excluding the bibliography, the exploration is at the upper end of the usual guidance and appears close to single spacing in places. Quality, not page count, is decisive, but repetition makes a 4 in criterion A less secure.	Merge the raw-data tables, shorten elementary algebra and replace three near-identical model graphs with one comparison figure. Aim for about 15–17 content-rich pages.
5	Tables 1–1.9	Numbering alternates between “Table 1”, “Table 1.1”, “Table 1.2”, and so on. The decimal system implies sub-tables even where the items are independent.	Use **Table 1, Table 2, …** throughout, or use **Table 1a–1d** only for deliberately grouped data tables.
6	Figures 1–1.7	The prose refers to both “Figure 1.0” and “Figure 1”, while captions use another sequence.	Choose one continuous figure sequence and make every in-text reference match it exactly.
7	Research question	“for say 300 mL” is filler language and reduces precision.	**“How accurately does Newton’s law of cooling model the cooling of 300 mL of hibiscus tea, and what time does the model predict for the temperature to fall below $45^\circ C$ in a room at $22.4^\circ C$?”**

No.	Location	Error or risk, and why it matters	Exact correction or possible replacement wording
8	Research question	“How accurately” is not operationalised. It is unclear whether accuracy means RMSE, residual behaviour, threshold-time error or performance on a new trial.	Add: **“Accuracy will be assessed using RMSE in degrees Celsius, residual plots and error in the predicted $45^\circ C$ crossing time on a validation trial.”**
9	Introduction	“Temperature is continuous” is presented as an experimental fact, although measurements are discrete. Continuity is a modelling assumption.	**“I modelled temperature as a continuous function of time, although readings were taken at discrete five-minute intervals.”**
10	Introduction	“The surrounding room places a lower restriction” is physically too absolute. Evaporation can cool a liquid below the dry-bulb ambient temperature.	**“Under the simplified Newton cooling model, and neglecting evaporative cooling below ambient temperature, room temperature is the long-term equilibrium value.”**
11	Introduction	Subject–verb agreement is wrong in “The bright red colour and sour taste makes”.	**“The bright red colour and sour taste make the drink especially enjoyable.”**
12	Introduction	“thermal equillibrium” is misspelled.	Use **“thermal equilibrium.”**
13	Introduction	The claim that hibiscus tea is less viscous than coffee is unsupported and does not explain the chosen drinking temperature.	Remove it, or write: **“I chose $45^\circ C$ from repeated personal trials; it is a subjective usability threshold rather than a universal preference.”**
14	Background	The discussion of anthocyanin extraction is only weakly connected to the mathematical question and reduces concision.	Reduce it to: **“I used the same hot-infusion procedure in every trial so that preparation differences would not dominate the cooling data.”**
15	Methodology	The three measured room temperatures are not displayed as raw data, so the stated mean of $22.4^\circ C$ is not reproducible.	Give the initial, final and mean room reading for each trial, or show the ambient time series in a compact table.
16	Methodology	“A fluctuation of about $\pm 0.3^\circ C$” is ambiguous. It could mean measurement uncertainty, observed range or deviation from the mean.	**“During the trials, the room reading ranged from … to … $\square^\circ C$; the thermometer resolution was … $\square^\circ C$.”**
17	Methodology	“Using three trials should eliminate random errors” is false. Repetition does not eliminate random error; it helps estimate variability and can reduce its influence on the mean.	**“Three repeated trials do not remove random error, but they allow variability to be estimated and reduce its influence on the mean.”**
18	Methodology	Display resolution and measurement uncertainty are treated as identical. A display reading to $0.1^\circ C$ does not prove an uncertainty of $\pm 0.1^\circ C$.	State the manufacturer’s accuracy or calibration result. Otherwise distinguish carefully between **resolution** and **estimated measurement uncertainty**.
19	Methodology	Timer resolution, reading delay and probe response time are omitted. These effects matter near the threshold.	State the timer resolution, typical reading delay and probe response time; discuss or test their likely effect on crossing-time estimates.
20	Raw data	Three almost identical tables occupy substantial space and make between-trial comparison awkward.	Use one table with columns T_1, T_2, T_3, mean and sample standard deviation.
21	Processed data	Only the mean is reported. Without a standard deviation or error bars, the reproducibility of the trials is invisible.	Calculate s_T at every time and show it in the table or as graph error bars.
22	Processed data	Averaging first and fitting only the mean curve hides variation in fitted parameters and threshold times.	Fit each trial separately, or at minimum calculate the $45^\circ C$ crossing time for each trial.
23	Figure 1	Colour alone distinguishes points and the asymptote, which is poor for greyscale printing and colour-vision accessibility.	Use different markers and line styles; do not describe series only as “red” and “black”.

No.	Location	Error or risk, and why it matters	Exact correction or possible replacement wording
24	Data/method	Humidity is described as uncontrollable but was not measured. Because evaporation is later used as an explanation, the claim remains untested.	Measure humidity, or identify evaporation only as one possible and unconfirmed explanation.

4.2 Mathematics, modelling and statistics

No.	Location	Error or risk, and why it matters	Exact correction or possible replacement wording
25	After Figure 1	“A horizontal asymptote at $t=22.4$” confuses time with temperature.	**“A horizontal asymptote at $T=22.4^{\circ}C$ was added.”**
26	After Figure 1	Boyle’s law does not explain the asymptote. Boyle’s law relates gas pressure and volume at constant temperature.	**“In the simplified Newton cooling model, net heat transfer tends to zero as T approaches the ambient temperature T_a.”**
27	After Figure 1	“A substance cannot decrease its temperature below that of its environment” is too absolute and ignores evaporative cooling.	**“Without an additional cooling mechanism such as evaporation, the Newton model approaches T_a asymptotically and does not cross it.”**
28	After Figure 1	“Negative correlation” is imprecise for a curved time series, and no correlation coefficient is calculated.	**“The data show a monotonic, nonlinear decrease in temperature with time.”**
29	Attempt A	Beginning with $T(t)=e^{-t}$ is dimensionally incomplete: the right side is dimensionless while T has temperature units.	Write the shape first as $f(t)=e^{-t}$; after scaling, use $T(t)=T_a+Ae^{-kt}$ with A in $\square^{\circ}C$ and k in min^{-1}.
30	Attempt A	k is called a “horizontal stretch factor”. In e^{-kt}, k is a rate and $1/k$ is the characteristic time scale.	**“The cooling-rate constant k has units min^{-1}; the characteristic time is $\tau=1/k$.”**
31	Attempt A	m is treated as a vertical scale factor but no unit is given.	Use A and define it: **“$A=T(0)-T_a$ is the initial excess temperature, measured in $\square^{\circ}C$.”**
32	Attempt A	The assertion that the first and last observations “should create the most representative model” is unsupported. Endpoints may contain error and all intermediate readings are ignored.	**“This endpoint estimate captures the total observed change, but it is sensitive to both endpoint readings and is used only as a preliminary benchmark.”**
33	Determining constants A	“Utmost accuracy” and six reported decimal places are false precision when temperatures are rounded to $0.1^{\circ}C$.	Retain unrounded values internally, but report, for example, $k=0.0473\,min^{-1}$ and explain the calculation precision.
34	Logarithm steps	$\ln(T-T_a)$ formally takes the logarithm of a dimensional quantity. Saying that logarithms have no units does not resolve this.	Define $\theta=(T-T_a)/(1^{\circ}C)$ and use $\ln\theta$. At SL, it is also acceptable to explain clearly that numerical values of the temperature difference measured in degrees Celsius are being transformed.
35	Attempt B	The symbol m already denotes exponential amplitude and is then reused as the slope in $y=mx+c$.	Use $y=bt+c$, where $b=-k$ and $c=\ln A$.
36	Attempt B	The line “$mx=kt$” has the wrong sign and is not a correct variable mapping.	Write $y=\ln(T-T_a)$, $x=t$, $b=-k$ and $c=\ln A$, so $y=bx+c$.
37	Table 1.5	The heading is “Time (s)” even though every observation and calculation uses minutes.	Change it to **“Time (min)”.**

No.	Location	Error or risk, and why it matters	Exact correction or possible replacement wording
38	Attempt B	The claim that changing T_a for each row would mean the equation no longer describes a function is mathematically false. A time-dependent ambient temperature still defines a function, but not this simple linearisation.	**“A time-dependent ambient temperature would produce a non-autonomous differential equation and invalidate this simple linearisation.”**
39	Figure 1.3	“98.71% of the variation in temperature is explained by time” misinterprets R^2. The regression response is $\ln\left(T-T_a\right)$, not T.	**“$R^2=0.9870$ means that 98.70% of the variation in the transformed response $\ln\left(T-T_a\right)$ is accounted for by its linear relation with time in this data set.”**
40	Figure 1.3	“a exponential equation” is grammatically incorrect.	Use **“an exponential equation.”**
41	Regression B	The graph shows approximately $y=-0.05100x+4.20476$, while the prose uses 4.20430 and $A=66.9735$. The rounded visible values give $A=e^{4.204761}\approx 67.0046$.	Use one unrounded regression output. From the visible means: $T(t)=22.4+67.0046e^{-0.0509973t}$.
42	Equation B/Table 1.7	The Table 1.7 predictions match the graph version of B with $A\approx 67.0046$, not the printed equation B with $A=66.9735$. The equation, table and RMSE therefore cannot all be reproduced from one another.	Generate the table and RMSE automatically from the same final spreadsheet formula or equation; do not retype parameters in several places.
43	Tables 1.7/parameters	Four to six decimal places in predicted temperatures imply accuracy unsupported by the instrument or model.	Keep more digits internally, but report approximately $0.1^\circ C$ or $0.01^\circ C$ in the text and state the uncertainty.
44	After Figure 1.2	A “repeated pattern of residuals” is asserted without a residual plot. Criteria D and E require evidence for this claim.	Plot $r_i=T_i-\hat{T}_i$ against t_i for each serious candidate, add a zero line and interpret the pattern.
45	After Figure 1.2	Air movement, thermal mass and evaporation are presented as causes without being measured or separated.	**“Possible explanations include …, but the present data cannot identify which mechanism caused the residual pattern.”**
46	Newton’s law	The unit of k is omitted, so the differential equation is not dimensionally documented.	Define $k>0$ in min^{-1}, T and T_a in ❑$^\circ C$, and t in minutes.
47	Differential equation	The derivation moves from $\ln\left\|T-T_a\right\|=-kt+C$ to $T-T_a=e^Ce^{-kt}$ without explaining why the absolute value may be removed.	**“All measured temperatures satisfy $T>T_a$, so $T-T_a$ is positive and the absolute value may be removed on the observed domain.”**
48	Differential-equation solution	A new parameter Z is introduced even though its meaning follows immediately from the initial condition; it is also left without a unit.	Use $A=T(0)-T_a$. Then $T(t)=T_a+\left(T_0-T_a\right)e^{-kt}$.
49	Attempt C	The justification for selecting $t=15$ and $t=45$ is weak. $T(45)=28.2^\circ C$ is not close to the drinking threshold of $45^\circ C$.	If a threshold-local two-point estimate is required, use the 20- and 25-minute observations. Prefer fitting all selected data by nonlinear regression.
50	Attempt C	The two points were selected after looking at the data. This creates selection bias and can make C look artificially good in a chosen interval.	State the selection rule before analysis, or use separate training and validation data.

No.	Location	Error or risk, and why it matters	Exact correction or possible replacement wording
51	Equation C	The printed equation is $T(t)=78.7149e^{-0.57955t}+22.4$, but the calculations use $-0.057955t$. The printed decay rate is ten times too large.	Use $T_C(t)=22.4+78.7149e^{-0.057955t}$ consistently, then regenerate every table and graph.
52	Figure 1.5	The horizontal axis reads “Time (seconds)” although the numerical values are minutes. A sentence below the graph does not repair the incorrect axis.	Correct the graph itself to **“Time (minutes)”.**
53	Models A–C	A, B and C are not three different physical models. All have the form $T=T_a+Ae^{-kt}$; only the parameter-estimation method changes.	Retitle the section **“Comparing three parameter-estimation methods for the same Newton cooling model.”**
54	Modelling method	Because RMSE in degrees Celsius is the chosen criterion, the most direct method is missing: nonlinear least squares in the original temperature scale.	Minimise $SSE(A,k)=\sum_i \left[T_i-\left(T_a+Ae^{-kt_i}\right)\right]^2$ and compare this result with the log-linear estimate.
55	RMSE section	RMSE is not generally “the standard deviation of the residuals”. That interpretation requires additional assumptions and a different denominator or centring convention.	**“RMSE is the square root of the mean squared residual and measures typical prediction error in the original temperature units.”**
56	RMSE section	“A lower result shows that the regression line is a better suit” is grammatically and mathematically imprecise; not all candidates are regression lines.	**“When models are evaluated on the same observations, a lower RMSE indicates a closer fit in degrees Celsius.”**
57	RMSE calculation	No worked RMSE example or unambiguous spreadsheet formula is supplied. This lack of reproducibility allowed the incorrect adjusted value to pass unnoticed.	Include a residual table or one fully substituted calculation, and name the software and formula used.
58	Table 1.9	The stated $1.991^{\circ}C$ for C on $t\geq 10$ cannot be reproduced from the displayed values. Table 1.7 gives about $2.050^{\circ}C$.	Recalculate: $\sqrt{\frac{1}{11}\sum_{t=10}^{60}\left(T_i-\hat{T}_i\right)^2}\approx 2.050^{\circ}C$.
59	Adjusted comparison	C is evaluated on $t\geq 10$, while A and B are quoted on $t\geq 0$. RMSE values from different observation sets cannot be compared directly.	Evaluate every equation on the same domain. From the visible data, for $t\geq 10$: A 3.243, B 1.604 and C $2.050\,\square^{\circ}C$.
60	Selection of C	Under a fair comparison, B—not C—is best among A, B and C for $t\geq 10$. The central choice in the exploration is therefore not supported by its own data.	Select B, or preferably use a direct nonlinear fit and test it on a separate trial.
61	Interpretation of C	C’s high initial-temperature estimate is mainly caused by fitting two observations and extrapolating backwards, not by holding $T_a=22.4^{\circ}C$ constant.	**“The unrealistic intercept results from fitting only the 15- and 45-minute observations and extrapolating that curve back to $t=0$.”**
62	Interpretation	Warming of the local air and heat returned by the cup are described as established explanations, but the data do not separate these mechanisms.	Label them as hypotheses. A coupled cup–liquid model may be mentioned as an extension, not as a demonstrated cause.
63	Removing 0–10 min	“I would never drink the tea immediately” does not automatically justify deleting early data because the requested time is still measured from $t=0$.	Define the relevant fitting domain before analysis, retain the original time origin and fit and test every method on the same domain.
64	Validation	The same data are used both to choose or fit the model and to assess it. This measures in-sample fit, not independent predictive accuracy.	Fit with Trials 1 and 2 and validate on Trial 3, then rotate the held-out trial, or collect a fourth independent validation trial.

4.3 Conclusion, reflection, sources and language

No.	Location	Error or risk, and why it matters	Exact correction or possible replacement wording
65	Answering the question	The model value of 21.53 min and the linearly interpolated value of 24.00 min differ by 2.47 min. “Reasonably close” is too vague without quantification.	**“The model predicts 21.53 min, whereas interpolation between the adjacent observations gives 24.00 min; the model predicts the crossing 2.47 min earlier.”**
66	Conclusion	“Maximum amount of time” sounds more exact than the data and model justify.	**“The selected model gives an estimated crossing time of about 22 minutes; the five-minute observations place the crossing between 20 and 25 minutes.”**
67	Conclusion	“If I begin drinking within twenty minutes, the model predicts the tea will remain above …” is ambiguous: above the threshold at the beginning, or throughout drinking?	**“The model predicts that the tea is still above $45^{\circ}C$ at $t=20$ min.”**
68	Result	No uncertainty interval is given for the threshold time. One decimal value is not adequate with five-minute sampling and visible model bias.	Report the trial-by-trial interpolations: approximately 24.15, 23.75 and 24.10 min; mean 24.00 min and sample SD about 0.22 min. Discuss model uncertainty separately.
69	Limitations	“More than three trials … permit uncertainty intervals” is false. Three trials already permit a preliminary spread estimate; more trials improve its reliability.	**“Three trials allow a preliminary estimate of variability; additional trials would make an uncertainty interval more reliable.”**
70	Limitations	“Recording every minute” is reasonable but not targeted to the research question.	Measure every 30–60 seconds between about 18 and 28 minutes; longer intervals may be sufficient elsewhere.
71	Limitations	Measuring mass instead of volume is suggested because of evaporation, but one initial mass would not show evaporative loss.	Weigh cup and tea during the experiment, or compare initial and final mass without materially altering heat transfer.
72	Limitations	Making T_a a function of time is mathematically promising, but the resulting equation is not stated, so the suggestion remains decorative.	Either test $T_a=22.1, 22.4, 22.7^{\circ}C$ in a sensitivity analysis or discuss $\frac{dT}{dt}=-k\left[T-T_a(t)\right]$ and an appropriate solution method.
73	Limitations	A magnetic stirrer would improve temperature homogeneity but also alter convection and therefore k.	Treat stirring as a new controlled design, not as a neutral correction to the existing trials.
74	Reflection	The short “Reflection” section largely summarises methods. The strongest reflection is scattered elsewhere.	Use the section to answer a mathematical question: why did a high linearised R^2 not produce the lowest temperature RMSE, and how did that change the model choice?
75	Assessment-style tone	“Since both equations were created mainly from the shape … I completed additional research” sounds like a retrospective rubric narrative rather than a genuine mathematical decision.	**“The residual pattern suggested that a fitted exponential alone did not explain why this form should arise, so I derived the model from a rate equation.”**
76	Assessment-style tone	“The mathematics developed from a visual conjecture into a physically motivated model” echoes assessment language rather than a personal thought process.	**“At first I chose the exponential form from the graph. Deriving Newton’s law later showed which assumptions were hidden in that choice.”**
77	Assessment-style tone	Stock transitions such as “This initially suggests …”, “On closer observation …” and “Nevertheless, it answers the investigation more directly” sound formulaic, especially when the numerical decision does not follow.	State the numbers directly: **“On $t\geq 10$, model B has RMSE $1.604^{\circ}C$ and model C $2.050^{\circ}C$; therefore B fits this domain more closely.”**

No.	Location	Error or risk, and why it matters	Exact correction or possible replacement wording
78	Assessment-style tone	"The investigation has transformed an ordinary routine into a practical use of …" is a generic closing sentence.	**"For my preparation method, the data place the $45^{\circ}C$ crossing between 20 and 25 minutes. The remaining uncertainty is large enough that I would use 20 minutes as a cautious reminder."**
79	Sources	Lienhard, Khan Academy and Abraham et al. appear in the bibliography but are not clearly cited in the body. This obstructs traceability and may look like bibliography padding.	Cite every source at the exact claim it supports, and remove unused sources.
80	Introduction image	The hibiscus image has no visible credit. The statement about experimental temperatures and graphs does not cover a third-party image.	Remove the image, or give creator, URL and licence in the caption and bibliography.
81	Works Cited	Citation style, DOI/URL format and access dates are inconsistent.	Apply one style throughout, such as MLA 9 or APA 7; give DOIs as https://doi.org/... where available.
82	OpenStax source	The listed OpenStax address "4-key-equations" does not clearly identify the Newton-cooling section being discussed.	Verify the exact section title and cite the direct address.
83	Reproducibility	"Spreadsheet and graphing software" does not name the program, version, regression option or formula range.	State the software and method, for example: **"I regressed $\ln(T-T_a)$ on t in … and calculated RMSE using …"** Explain the central calculation in the text as well.
84	Language	Repeated errors reduce the professional impression: "seperated", "linearisaton", "Newtons law", "degrees celsius", missing articles and inconsistent capitalisation.	Use **separated, linearisation, Newton's law, degrees Celsius**, then complete a full spelling and grammar check.
85	Academic integrity	If AI-generated wording or calculations are included in the submitted exploration, current IB guidance requires them to be credited in the body and appropriately referenced. Uncredited AI-generated material is not the student's own work.	Understand and verify every mathematical idea, then rewrite it in the student's own voice. Disclose any retained AI-generated material according to the school's and IB's requirements; do not submit this review as personal work.

5. Essential mathematical rebuild

5.1 One model, several parameter-estimation methods

A, B and C should not be presented as different physical models. Their common assumption is Newton's law of cooling:

$$\frac{dT}{dt}=-k(T-T_a), k>0,$$

with solution

$$T(t)=T_a+Ae^{-kt}, A=T(0)-T_a.$$

Only the method used to estimate A and k changes:

1. an endpoint estimate;
2. linear regression after logarithmic transformation;
3. a two-point estimate using observations chosen after inspection;
4. direct nonlinear regression in the original temperature scale.

The fourth method is aligned with an RMSE criterion in degrees Celsius because it directly minimises

$$SSE(A,k)=\sum_{i=1}^{n}\left[T_i-\left(T_a+Ae^{-kt_i}\right)\right]^2.$$

5.2 Independent calculation from the displayed mean data

With $T_a=22.4^{\circ}C$ fixed, direct nonlinear least squares on the visible mean values gives approximately

$$\hat{T}(t)=22.4+58.87\,e^{-0.04291t},$$

where t is measured in minutes. This yields

$$RMSE\approx 1.819^{\circ}C$$

and

$$t_{45}=-\frac{1}{0.04291}\ln\left(\frac{45-22.4}{58.87}\right)\approx 22.31\,min.$$

These values are an independent check based on the rounded numbers visible in the PDF. They should be recalculated from the original unrounded observations and understood by the student before use.

5.3 Fair comparison of the displayed equations

Method	RMSE, all observations	RMSE for $t\geq 10$ min	Predicted t_{45}
A: endpoint estimate	$3.130^{\circ}C$	$3.243^{\circ}C$	19.28 min
B: corrected log regression	$3.530^{\circ}C$	$1.604^{\circ}C$	21.30 min
C: 15/45-minute points	$7.214^{\circ}C$	$2.050^{\circ}C$	21.53 min
Direct nonlinear regression	$1.819^{\circ}C$	$1.805^{\circ}C$	22.31 min

C is neither the best global estimate nor the best estimate for $t\geq 10$. The observed crossing time obtained by linear interpolation between 20 and 25 minutes is approximately

$$24.15,\,23.75,\,24.10\,min$$

for the three trials, with mean about 24.00 min and sample standard deviation about 0.22 min. The difference between 24.00 min and the direct nonlinear prediction of 22.31 min is about 1.69 min. This is valuable reflection: between-trial variability is small, yet the simple one-exponential model has a systematic predictive discrepancy.

6. Revision plan for a top mark

Priority 1 — essential before submission

1. Correct the factor-of-ten error in equation C and every seconds/minutes error.
2. Recreate equation B from one parameter set; generate its table, graph and RMSE from the same source.
3. Remove Boyle’s law and explain the asymptote through Newton’s cooling model.
4. Compare all candidates on the same observations.
5. Correct the adjusted RMSE and rewrite the model choice.

Priority 2 — largest likely mark gain

1. Add direct nonlinear regression in the original temperature scale.
2. Add and interpret a residual plot for the preferred model.
3. Use Trial 3 as independent validation, rotate the held-out trial, or collect a fourth trial.
4. Report the threshold time for each trial and quantify its spread.
5. Add a notation-and-units table for t, T, T_a, A, k, r_i and RMSE.

Priority 3 — from strong to excellent

1. Reduce the exploration by about 3–5 pages without removing substantive mathematics.
2. Show personal engagement through consequential decisions, not repeated use of “I”.
3. Attach a quantitative consequence to each important limitation.
4. Remove unused sources and the uncredited image.
5. Standardise every figure, table, page number, unit and decimal place.

7. Natural replacement passages

These passages are deliberately direct and less rubric-like. They should be understood, checked against the student’s actual work and rewritten in the student’s own voice rather than copied mechanically.

Research question and measure of accuracy

Research question. How well does Newton’s law of cooling describe the temperature of 300 mL of hibiscus tea in a room at approximately $22.4^\circ C$, and what time does the model predict for the tea to fall below my chosen threshold of $45^\circ C$? I will assess the model using residuals, RMSE in degrees Celsius and its error when predicting a trial that was not used to estimate the parameters.

Physical modelling assumption

I treated room temperature as constant and assumed that the net cooling rate was proportional to the tea’s temperature difference from the room. With $T_a = 22.4^\circ C$ and $k > 0$, this gives $\frac{dT}{dt} = -k(T - T_a)$. Because every recorded tea temperature was above T_a, solving the equation gives $T(t) = T_a + Ae^{-kt}$, where A is the initial excess temperature and k is measured in min^{-1}.

Model fitting

Linearising the data is useful for checking whether an exponential pattern is plausible, but it changes the weighting of errors. Since my final accuracy measure is RMSE in degrees Celsius, I also fitted $T(t) = 22.4 + Ae^{-kt}$ directly by minimising squared residuals in the original temperature scale. This avoids choosing two convenient observations and uses every value in the selected training set.

Fair comparison

I evaluated every parameter-estimation method on the same observations. On the restricted domain $t \geq 10$ min, the RMSE values for A, B and C were approximately $3.243^\circ C$, $1.604^\circ C$ and $2.050^\circ C$, respectively. C was therefore not the most accurate of these three estimates on the domain relevant to the threshold. The direct nonlinear fit gave the lowest full-data RMSE, but its residuals still showed a systematic pattern, so the one-exponential model should not be treated as exact.

Conclusion

The direct nonlinear model predicts that the tea reaches $45^\circ C$ after about 22.3 minutes. Linear interpolation of the measurements surrounding the threshold gives about 24.0 minutes. The difference of roughly 1.7

minutes is larger than the variation between the three interpolated trial values, suggesting systematic model error rather than only random measurement variation. I would therefore report a practical crossing window of 22–24 minutes, while the present five-minute sampling establishes only that the crossing occurred between 20 and 25 minutes.

Critical reflection

The high R^2 value of the linearised graph initially made the exponential model appear stronger than it was. However, R^2 described the transformed variable $\ln\left(T - T_a\right)$, whereas my practical question concerns errors in degrees Celsius and minutes. Refitting the model in the original scale changed the preferred parameter estimates. A visually straight transformed graph is therefore evidence for a model form, but it is not proof that its predictions are most accurate for the quantity I care about.

8. Assessment sources

1. International Baccalaureate, *Mathematics: analysis and approaches subject brief — first assessments 2021*.
2. International Baccalaureate, *Mathematics: analysis and approaches updates*, including the new course timeline and May 2029 first assessment.
3. International Baccalaureate, *Artificial intelligence in learning, teaching, and assessment*.

Final assessment

The investigation has strong potential. The decisive improvements are consistency, fair validation and honest quantitative criticism of the model—not additional pages or more complicated mathematics. If the five central mathematical errors are corrected and direct regression, residual analysis and validation are integrated coherently, the exploration can become a convincing top-band submission.