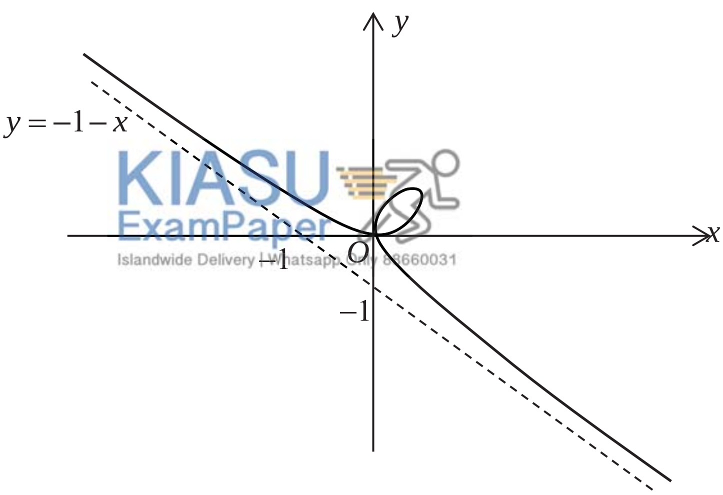
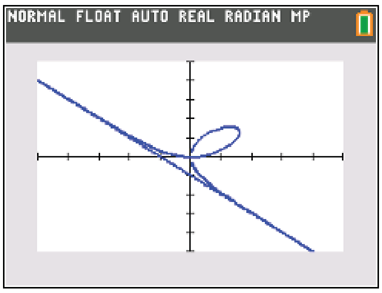
2019 JC2 H2 Prelim Exam P2 Markers Report

1(i)	$k(3\mathbf{a}+5\mathbf{b})=\mathbf{a}+\lambda(\mathbf{b}-\mathbf{a})$ Comparing coefficient of **a** and **b**, $\left.\begin{array}{l}3k=1-\lambda\\ 5k=\lambda\end{array}\right\}\therefore \lambda=\frac{5}{8},\ k=\frac{1}{8}$ $\therefore$p.v of the point of intersection is $\frac{3}{8}\mathbf{a}+\frac{5}{8}\mathbf{b}$.	Some made the mistake of using vector AB as the line equation Line AB is $r=a+\lambda(b-a)$ not $r=b-a$ Some assumed that the intersection point is R, making the mistake of equating $(3\mathbf{a}+5\mathbf{b})=\mathbf{a}+\lambda(\mathbf{b}-\mathbf{a})$ when it should have been $k(3\mathbf{a}+5\mathbf{b})=\mathbf{a}+\lambda(\mathbf{b}-\mathbf{a})$
(ii)	$\mathbf{a}\cdot\mathbf{b}=\lvert\mathbf{a}\rvert\lvert\mathbf{b}\rvert\cos 45^{o}=\frac{\lvert\mathbf{a}\rvert}{\sqrt{2}}$ $\lvert\overrightarrow{OR}\cdot\hat{\mathbf{b}}\rvert=\lvert(3\mathbf{a}+5\mathbf{b})\cdot\mathbf{b}\rvert=\lvert 3\mathbf{a}\cdot\mathbf{b}+5\mathbf{b}\cdot\mathbf{b}\rvert$ $=(3\mathbf{a}\cdot\mathbf{b}+5\lvert\mathbf{b}\rvert^{2})=\frac{3\lvert\mathbf{a}\rvert}{\sqrt{2}}+5$	Some wrote the dot product correctly in (i) but didn't put it to use in any way.
2	$\frac{1}{r(r-2)}-\frac{1}{r(r+2)}=\frac{(r+2)-(r-2)}{(r-2)r(r+2)}=\frac{4}{(r-2)r(r+2)}$ $\sum_{r=3}^{n}\frac{1}{(r-2)r(r+2)}=\frac{1}{4}\sum_{r=3}^{n}\left(\frac{1}{r(r-2)}-\frac{1}{r(r+2)}\right)$ $=\frac{1}{4}\left(\frac{1}{3(1)}-\cancel{\frac{1}{3(5)}}\right.$ $+\frac{1}{4(2)}-\cancel{\frac{1}{4(6)}}$ $+\cancel{\frac{1}{5(3)}}-\cancel{\frac{1}{5(7)}}$ $+\cdots$ $+\cancel{\frac{1}{(n-2)(n-4)}}-\cancel{\frac{1}{(n-2)n}}$ $+\cancel{\frac{1}{(n-1)(n-3)}}-\frac{1}{(n-1)(n+1)}$ $\left.+\cancel{\frac{1}{n(n-2)}}-\frac{1}{n(n+2)}\right)$ $=\frac{1}{4}\left(\frac{1}{3}+\frac{1}{8}-\frac{1}{(n-1)(n+1)}-\frac{1}{n(n+2)}\right)$ $=\frac{11}{96}+\frac{-\frac{1}{4}}{(n-1)(n+1)}+\frac{-\frac{1}{4}}{n(n+2)}$ $a=\frac{11}{96},b=c=-\frac{1}{4}$	A minority of students did not consider the given expression. Out of those who did use the result from considering the given expression, ~10-20% of them used the result wrongly (multiplying by 4 instead of by ¼). MOD was done well generally, with minor slips.

(i)	From previous result, $\sum_{r=3}^{n}\frac{1}{(r-2)r(r+2)}=\frac{11}{96}-\frac{1}{4(n-1)(n+1)}-\frac{1}{4n(n+2)}$ As $n\to\infty$, $\frac{1}{(n-1)(n+1)}, \frac{1}{n(n+2)}\to 0$, therefore $\sum_{r=3}^{\infty}\frac{1}{(r-2)r(r+2)}=\frac{11}{96}$.	Done well, but the presentation was off for 20-30%. Some even wrote As $r\to\infty$, $\frac{1}{(r-2)r(r+2)}\to\frac{11}{96}$. Students are reminded to <u>answer the question</u>.
(ii)	$\sum_{r=5}^{n}\frac{1}{r(r+2)(r+4)}$ Replace r by $r-2$, $\sum_{r-2=5}^{r-2=n}\frac{1}{(r-2)r(r+2)}=\sum_{r=7}^{n+2}\frac{1}{(r-2)r(r+2)}$ $=\sum_{r=3}^{n+2}\frac{1}{(r-2)r(r+2)}-\sum_{r=3}^{6}\frac{1}{(r-2)r(r+2)}$ $=\left(\frac{11}{96}-\frac{1}{4(n+1)(n+3)}-\frac{1}{4(n+2)(n+4)}\right)$ $-\left(\frac{11}{96}-\frac{1}{4(6-1)(6+1)}-\frac{1}{4(6)(6+2)}\right)$ $=\frac{83}{6720}-\frac{1}{4(n+1)(n+3)}-\frac{1}{4(n+2)(n+4)}$	Poorly done. A variety of similar approaches can be applied, but there were many conceptual errors with regard to how to the general term interacts with the dummy variable in the summation. There was a small number of scripts where the working was filled with errors arriving at the right answer (or close to). Marks may not have been awarded if the terms used in the working were not equal at any juncture.
3(a)(i)	$y^2=b^2\left(1-\frac{x^2}{a^2}\right)=\frac{b^2(a^2-x^2)}{a^2}$ $A=4\int_0^a y\ \mathrm{d}x=4\int_0^a\sqrt{b^2\left(1-\frac{x^2}{a^2}\right)}\ \mathrm{d}x$ $=4\int_0^a\sqrt{\frac{b^2(a^2-x^2)}{a^2}}\ \mathrm{d}x=\frac{4b}{a}\int_0^a\sqrt{a^2-x^2}\ \mathrm{d}x.$ (shown) $A=\frac{4b}{a}\int_0^a\sqrt{a^2-x^2}\ \mathrm{d}x$ $=\frac{4b}{a}\int_0^{\frac{\pi}{2}}\sqrt{a^2-(a\sin\theta)^2}(a\cos\theta)\ \mathrm{d}\theta$ $=\frac{4b}{a}\int_0^{\frac{\pi}{2}}\sqrt{a^2(1-\sin^2\theta)}(a\cos\theta)\ \mathrm{d}\theta$ $=\frac{4b}{a}\int_0^{\frac{\pi}{2}}a^2\cos^2\theta\ \mathrm{d}\theta$ $=2ab\int_0^{\frac{\pi}{2}}2\cos^2\theta\ \mathrm{d}x=2ab\int_0^{\frac{\pi}{2}}1+\cos 2\theta\ \mathrm{d}\theta$ $=2ab\left[\theta+\frac{1}{2}\sin 2\theta\right]_0^{\frac{\pi}{2}}=2ab\left(\frac{\pi}{2}\right)=\pi ab$	For the 'show' part, it is a simple result that aims to test the students' conceptual understanding of the application of integration to find area under a curve. Students' presentation has to demonstrate that understanding. Generally very well done, though the usual mistakes were still present (not changing the limits, making errors substituting in $\frac{\mathrm{d}x}{\mathrm{d}\theta}$). Some students did not know how to proceed with the integrand thereafter (but only minority). Of those who applied the cosine double angle formula, there was a significant minority who integrated $\cos 2\theta$ wrongly.

(a)(ii)	$V=2\pi\int_0^a y^2\ \mathrm{d}x$ or $\pi\int_{-a}^a y^2\ \mathrm{d}x$ $=2\pi\int_0^a \frac{b^2(a^2-x^2)}{a^2}\ \mathrm{d}x=\frac{2\pi b^2}{a^2}\int_0^a a^2-x^2\ \mathrm{d}x$ $=\frac{2\pi b^2}{a^2}\left[a^2x-\frac{x^3}{3}\right]_0^a=\frac{2\pi b^2}{a^2}\left(\frac{2a^3}{3}\right)$ $=\frac{4\pi ab^2}{3}$	Poorly done by quite a number of students. These students may have been confused by the integral they showed in (i), not helped by their poor understanding of finding volume of revolution by applying integration. Simplify answers too!
(b)	$x^2+y^2=a^2$ $2x+2y\frac{\mathrm{d}y}{\mathrm{d}x}=0$ $\frac{\mathrm{d}y}{\mathrm{d}x}=-\frac{x}{y}$ $S=\int_a^b 2\pi y\sqrt{1+\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\ \mathrm{d}x$ $=\int_{-r}^r 2\pi y\sqrt{1+\left(-\frac{x}{y}\right)^2}\ \mathrm{d}x$ $=\int_{-r}^r 2\pi\sqrt{x^2+y^2}\ \mathrm{d}x$ $=\int_{-r}^r 2\pi\sqrt{x^2+(r^2-x^2)}\ \mathrm{d}x$ $=\int_{-r}^r 2\pi r\ \mathrm{d}x=2\pi r\ [x]_{-r}^r=4\pi r^2$ Alternative: $y=\sqrt{a^2-x^2}$ $\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{-x}{\sqrt{a^2-x^2}}$ $S=\int_a^b 2\pi y\sqrt{1+\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\ \mathrm{d}x$ $=\int_{-r}^r 2\pi\sqrt{r^2-x^2}\sqrt{1+\left(\frac{-x}{\sqrt{r^2-x^2}}\right)^2}\ \mathrm{d}x$ $=\int_{-r}^r 2\pi\sqrt{r^2-x^2}\sqrt{1+\frac{x^2}{r^2-x^2}}\ \mathrm{d}x$ $=\int_{-r}^r 2\pi\sqrt{(r^2-x^2)+x^2}\ \mathrm{d}x$ $=\int_{-r}^r 2\pi r\ \mathrm{d}x=2\pi r\ [x]_{-r}^r=4\pi r^2$	Students generally made good progress with this part, applying the formula with ease. A common conceptual error was integrating r with respect to x to obtain $\frac{r^2}{2}$.

4	When $a=1$, $x^3+y^3=3xy$. Let $y=xt$, then $x^3+(xt)^3=3x(xt)$ $\Rightarrow \quad x^3+x^3t^3=3x^2t$ $\Rightarrow \quad x^3(1+t^3)=3x^2t$ $\Rightarrow \quad x=\dfrac{3t}{1+t^3}$. (shown) For x to be defined, $1+t^3\neq 0$ Hence $t\neq -1$. Therefore $k=-1$ Since $y=xt$, hence $y=\left(\dfrac{3t}{1+t^3}\right)t=\dfrac{3t^2}{1+t^3}$.	The proof for x was generally well done. Working for find y in terms of t varied from a one-liner to a full page working when the question says "write down the expression…" which students should suspect that the answer is obvious. Some students neglected to write down the value of k.
(i)	$x+y=\dfrac{3t}{1+t^3}+\dfrac{3t^2}{1+t^3}$ $=\dfrac{3t(1+t)}{1+t^3}$ $=\dfrac{3t(1+t)}{(1+t)(t^2-t+1)}$ $=\dfrac{3t}{t^2-t+1}$. (shown) Oblique asymptote is when $x\to\pm\infty$. Since $x=\dfrac{3t}{1+t^3}$, $x\to\pm\infty$ when $t\to -1$. When $t\to -1$, $x+y=\dfrac{3t}{t^2-t+1}\to\dfrac{-3}{(-1)^2-(-1)+1}=\dfrac{-3}{3}=-1$. i.e. $x+y\to -1$ $\Rightarrow \quad y\to -1-x$. Therefore, the oblique asymptote of the curve is $y=-1-x$.	In this show question, students either did not know how to simplify $\dfrac{3t(1+t)}{1+t^3}$, or some just did $\dfrac{3t(1+t)}{1+t^3}=\dfrac{3t}{t^2-t+1}$ without showing the factorization. Students need to realise that nothing in a "show" or "prove" question should be assumed as obvious. Only a few students knew that oblique asymptote occurs when $x\to\pm\infty$, not when $t\to\pm\infty$.
(ii)		Graph was badly drawn. Students need to know how to interpret their G.C. sketch: 

		1. Why is there a "hole" in the graph near the origin? 2. Is that line that looks like a straight line part of the graph? Sketches that included a sharp turn to draw the straight line as part of the graph did not get any credit. Nor did sketches with a gap near the origin.
(iii)	Differentiating the cartesian equation implicitly, $3x^2+3y^2\frac{\mathrm{d}y}{\mathrm{d}x}=3x\frac{\mathrm{d}y}{\mathrm{d}x}+3y$ $\Rightarrow\ \left(y^2-x\right)\frac{\mathrm{d}y}{\mathrm{d}x}=y-x^2$ i.e. $\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{y-x^2}{y^2-x}$. When $t=2,\ x=\frac{2}{3},\ y=\frac{4}{3}$ and $\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\frac{4}{3}-\left(\frac{2}{3}\right)^2}{\left(\frac{4}{3}\right)^2-\frac{2}{3}}=\frac{4}{5}$. Hence equation of tangent when $t=2$: $\frac{y-\frac{4}{3}}{x-\frac{2}{3}}=\frac{4}{5}\ \Rightarrow\ y=\frac{4}{5}x+\frac{4}{5}$. If the tangent cuts the curve again, using the parametric equations of the curve to substitute into the equation of the tangent, $\frac{3t^2}{1+t^3}=\frac{4}{5}\left(\frac{3t}{1+t^3}\right)+\frac{4}{5}$ $\Rightarrow\ 3t^2=\frac{4}{5}(3t)+\frac{4}{5}\left(1+t^3\right)$ $\Rightarrow\ 15t^2=12t+4+4t^3$ $\Rightarrow\ 4t^3-15t^2+12t+4=0$ $\Rightarrow\ t=-\frac{1}{4}$, or $t=2$ (tangent here) Therefore the tangent cuts the curve again at $t=-\frac{1}{4}$, with coordinates $\left(-\frac{16}{21},\frac{4}{21}\right)$.	This part of the question required students to choose between the given Cartesian equation or the equivalent parametric equations to use to find $\frac{\mathrm{d}y}{\mathrm{d}x}$. Most students chose to differentiate the parametric equations, with many many remembering quotient rule wrongly. Some used product rule instead but also fumbled with the algebra. The most efficient method is to differentiate the Cartesian equation implicitly, then substituting $t=2$ to find *x* and *y* at the point to find gradient of the tangent. The intersection between tangent and original curve should be a familiar question. The parametric equations should be substituted into the equation of the tangent to solve for *t*. Technically, even without having done the preceding parts, this question is a standard one that could have easily been done with just the Cartesian equation and the G.C.
5(i)	Number of ways $={}^{12}C_4$ $=495$	Most common error was to use ${}^{12}P_4$ instead. But order in this case is already

<table>
<tr><td></td><td></td><td>predetermined by heights, so we just need to find how many ways a group of 4 can be chosen and there is only 1 way to arrange them by height.</td></tr>
<tr><td>5(ii)</td><td><u>Method 1</u>
Number of ways
$= 2^{12}-1$
$= 4095$

<u>Method 2</u>
Number of ways
$= {}^{12}C_1 + {}^{12}C_2 + {}^{12}C_3 ... + {}^{12}C_{12}$
$= 4095$</td><td>This question, unintentionally, has two interpretations:
1. Total number of ways to choose delegations with at least one student. This was the intended interpretation.
2. If n is assumed constant, then number of ways to choose n students from the team is just ${}^{12}C_n = \frac{12!}{n!(12-n)!}$. For this approach we need to see the formula in order to gain the second mark.</td></tr>
<tr><td>5(iii)</td><td>Probability
$= \frac{{}^{5}C_3 3! \times {}^{5}C_3 3! \ \times 2}{{}^{12}C_8 7!}$
$= \frac{2}{693}$ (or 0.00289)</td><td>This question proved a challenge for many with a majority neglecting to choose the number of boys and girls to be seated with Andy and Beth. The team has 12 members but only 8 are chosen to attend the dinner. Another problem is not knowing that the total number of ways is to count how many ways to choose any 8 to seat at a round table.</td></tr>
<tr><td>6(i)</td><td>$\text{P}(X=2) = 2\text{P}(\{1,3\}) + 2\text{P}(\{3,5\}) + 2\text{P}(\{5,7\})$
$= 2\left(\frac{1}{4}\right)\left(\frac{1}{3}\right) + 2\left(\frac{1}{4}\right)\left(\frac{1}{3}\right) + 2\left(\frac{1}{4}\right)\left(\frac{1}{3}\right)$
$= \frac{1}{2}$ (shown)
$\text{P}(X=4) = 2\text{P}(\{1,5\}) + 2\text{P}(\{3,7\})$
$= 2\left(\frac{1}{4}\right)\left(\frac{1}{3}\right) + 2\left(\frac{1}{4}\right)\left(\frac{1}{3}\right) = \frac{1}{3}$
<table><tr><td>x</td><td>2</td><td>4</td><td>6</td></tr><tr><td>$\text{P}(X=x)$</td><td>$\frac{1}{2}$</td><td>$\frac{1}{3}$</td><td>$\frac{1}{6}$</td></tr></table></td><td>(i) Unclear presentation for P(X=2) = 0.5 for many students.

Some wrote $6\left(\frac{1}{4}\right)\left(\frac{1}{3}\right) = \frac{1}{2}$ or $2\left(\frac{1}{4}\right)\left(\frac{1}{3}\right) + 4\left(\frac{1}{4}\right)\left(\frac{1}{3}\right)$ without explaining what these numbers meant.</td></tr>
<tr><td>6(ii)</td><td>Similarly
<table><tr><td>y</td><td>2</td><td>4</td><td>6</td></tr><tr><td>$\text{P}(Y=y)$</td><td>$\frac{1}{2}$</td><td>$\frac{1}{3}$</td><td>$\frac{1}{6}$</td></tr></table></td><td>(ii) some made the mistake
$\text{E}(W) = 2\left(\frac{1}{2}\right) + 4\left(\frac{1}{3}\right) + 6\left(\frac{1}{6}\right)$</td></tr>
</table>

<table>
<tr>
<td></td>
<td>P(player wins a prize)
$= \mathrm{P}(X=2)\mathrm{P}(Y=2)+\mathrm{P}(X=4)\mathrm{P}(Y=4)+\mathrm{P}(X=6)\mathrm{P}(Y=6)$
$=\left(\frac{1}{2}\right)^2+\left(\frac{1}{3}\right)^2+\left(\frac{1}{6}\right)^2$
$=\frac{7}{18}$</td>
<td>when it should have been
$\mathrm{E}(W)=2\left(\frac{1}{2}\right)^2+4\left(\frac{1}{3}\right)^2+6\left(\frac{1}{6}\right)^2$</td>
</tr>
<tr>
<td>6(ii)</td>
<td>Let W be the winnings per game.
$$\begin{array}{|c|c|c|c|c|}\hline w & 0 & 2 & 4 & 6 \\ \hline \mathrm{P}(W=w) & \frac{11}{18} & \left(\frac{1}{2}\right)^2 & \left(\frac{1}{3}\right)^2 & \left(\frac{1}{6}\right)^2 \\ \hline \end{array}$$
$\mathrm{E}(W)=2\left(\frac{1}{2}\right)^2+4\left(\frac{1}{3}\right)^2+6\left(\frac{1}{6}\right)^2=\frac{10}{9}$
Since the expected winnings of player $=\$\frac{10}{9} > \$1$, the game is to be played using $2 coupon in order to have profit per game.
$k=2$</td>
<td>Some didn’t realise that to win a prize you need P(X=Y), thus making the mistake of calculating E(X) = 10/3

Some students did not realise that amount won is associated with their different probabilities, they made the mistake of using prob $\frac{7}{18}$ with W (winning per game).

Some found E(W) $=\frac{10}{9}$ but made wrong conclusion for k = $1</td>
</tr>
<tr>
<td>7(i)</td>
<td>$\bar{x}=\frac{\sum(x-200)}{50}+200=209$

$s^2=\frac{1}{49}\left[55000-\frac{(450)^2}{50}\right]=\frac{50950}{49}$ or 1039.8 (5 s.f.)
Let μ be the population mean price of all concert tickets.

To test $\mathrm{H}_0:\mu=200$
against $\mathrm{H}_1:\mu\neq 200$ at the 4% level of significance.
Under H_0, $Z=\frac{\bar{X}-200}{s/\sqrt{50}}\sim\mathrm{N}(0,1)$ or $\bar{X}\sim\mathrm{N}\left(200,\ \frac{50950}{49(50)}\right)$
approximately by central limit theorem since sample size, 50, is large.

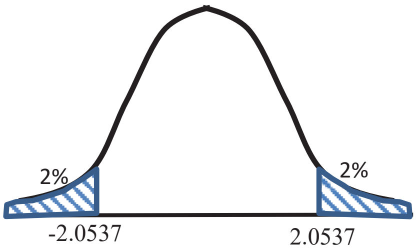
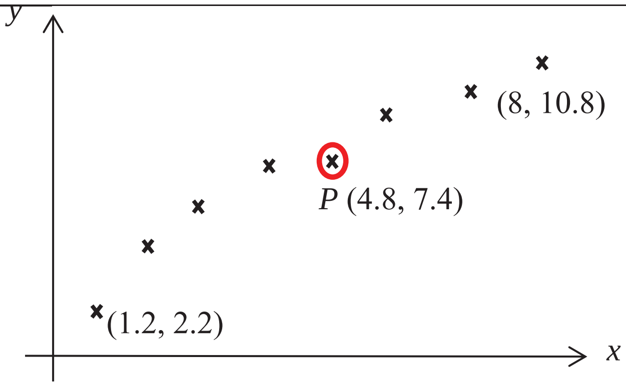
Value of test statistic $z=1.97$
p-value = 0.0484 > 0.04 (do not reject H_0)

There is insufficient evidence at the 4% level of significance that the mean concert ticket price is not $200.</td>
<td>Most students are able to present first step, i.e.
$\mathrm{H}_0:\mu=200$
$\mathrm{H}_1:\mu\neq 200$ at 4% level of sig.
Many DID NOT define μ.

Some students made the mistake of putting the wrong population mean 209 in place of the value 200 in the distribution,
$\bar{X}\sim\mathrm{N}\left(200,\ \frac{50950}{49(50)}\right)$

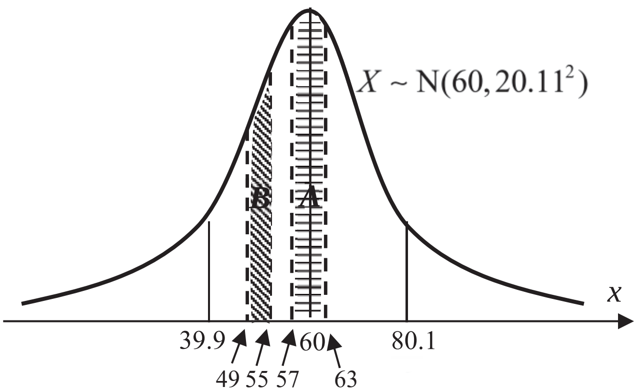
Some forgot to quote central limit theorem.

Phrasing of the conclusion was contradicting for some students, even though they knew it was ‘do not reject H_0’.</td>
</tr>
<tr>
<td>(ii)</td>
<td>Now given the standard deviation is 32.25
Under H_0, $\bar{X}\sim\mathrm{N}\left(200,\ \frac{32.25^2}{n}\right)$ by CLT since n is large</td>
<td>Some students misread the question.</td>
</tr>
</table>

	 For H_0 to not be rejected, the test statistic, z, must not be in critical region. Hence $-2.0537 < z < 2.0537$ i.e. $-2.0537 < \dfrac{206-200}{\frac{32.25}{\sqrt{n}}} < 2.0537$ i.e. $-2.0537 < \dfrac{6\sqrt{n}}{32.25} < 2.0537$ i.e. $-11.039 < \sqrt{n} < 11.039$ Hence largest value of n is 121.	(ii) Some made the mistake to assuming it is upper tail, writing $\dfrac{206-200}{\frac{32.25}{\sqrt{n}}} < 2.0537$ instead of $-2.0537 < \dfrac{206-200}{\frac{32.25}{\sqrt{n}}} < 2.0537$ Some also made the mistake of confusing it with (i) information, writing it as $-2.0537 < \dfrac{209-206}{\frac{32.25}{\sqrt{n}}} < 2.0537$
8(i)		Students should be using a cross 'x' to make the points instead of a small square box (as seem on the calculator screen).
(ii)	The scatter diagram suggests that as x increases, y increases at a decreasing rate, which is consistent with the model $y = c\ln x + d$ where c and d are positive. The model $y = ax^2 + b$ where a and b are positive suggests instead that as x increases, y increases at an increasing rate, which is not suitable for the data given. $r = 0.9927$ (4dp) 	Quite a number of students said that the points in the scatter diagram were distributed along a line, hence $y = c\ln x + d$ was the appropriate model (incorrectly assuming ln x is a **lin**ear function?) Some students mentioned the turning point in $y = ax^2 + b$. This is not relevant to the current question as we are only looking at a limited range of values of x. Some students also calculated r for both models before deciding $y = c\ln x + d$ was the better model as its value of r was closer to 1. This was

		not accepted as it did not make use of the scatter diagram. Finally, many students did not give r to 4 decimal places as stated in the question.
(iii)	Least squares regression line is $y = 1.3755 + 4.4612 \ln x$ (5 s.f.) $= 1.38 + 4.46 \ln x$ (3 s.f.)	A large number of students gave $y = 4.46 + 1.38 \ln x$, or otherwise chose the wrong model in part (ii) and hence lost marks here. Most students chose P correctly; some students thought it was the bottom left point (1.2, 2.2).
(iv)	When $y = 8$, $8 = 1.3755 + 4.4612 \ln x$ Hence $\ln x = \dfrac{8 - 1.3755}{4.4612}$ $\Rightarrow x = 4.42$ (3 s.f.) This estimate is reliable as $r_{y,\ln x}$ is very close to +1 and $y = 8$ is within the data range of y ($2.2 \le y \le 10.8$) used to obtain the regression line of y on ln$x$. The popularity index (variable x) is the independent variable in this context, therefore it is not appropriate to use the ln x on y or the x^2 on y regression lines to estimate popularity index x. 	Some students did not use the regression line in (iii), but recalculated a new regression line based on a linear model. Some students left their answer as $x = \dfrac{8 - 1.3755}{4.4612} = 1.48$. Many students omitted to mention that r is close to 1. Some students wrote that it was unreliable as the popularity index of a new artiste might not be accurate. This was not accepted. Finally, a large number of students stated the two models were inappropriate for various reasons – such as the fact that y could not be negative, or that the graph did not have a turning point. This reflects a misconception that the regression model can be extrapolated to values of x beyond those of the data.
9(i)	Let A be the event that the older twin Albert is late for practice, and B be the event that the younger twin Benny is late for practice. Given: $\text{P}(A) = 0.65$, $\text{P}(A \mid B) = 0.975$, $\text{P}(A \mid B') = 0.56875$.	Majority of students were able to do this. Those who could not wrongly interpreted the

	$\mathrm{P}(A\mid B)=0.975 \Rightarrow \dfrac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}=0.975$ $\Rightarrow \mathrm{P}(A\cap B)=0.975\,\mathrm{P}(B)$ $\mathrm{P}(A\mid B')=0.56875 \Rightarrow \dfrac{\mathrm{P}(A\cap B')}{\mathrm{P}(B')}=0.56875$ $\Rightarrow \mathrm{P}(A\cap B')=0.56875\,\mathrm{P}(B')$ Now, $\mathrm{P}(A)=\mathrm{P}(A\cap B)+\mathrm{P}(A\cap B')$ $0.975\mathrm{P}(B)+0.56875[1-\mathrm{P}(B)]=0.65$ $\mathrm{P}(B)=0.2$ (shown)	probability of 0.975 as $\mathrm{P}(A\cap B)$ instead of $\mathrm{P}(A\mid B)$. Some students had difficulty combining $\mathrm{P}(A\cap B)$ and $\mathrm{P}(A\cap B')$, not realizing that drawing a simple venn diagram would clearly lead to $\mathrm{P}(A)=\mathrm{P}(A\cap B)+\mathrm{P}(A\cap B')$.
9(ii)	$\dfrac{\mathrm{P}(A\cap B)}{0.2}=0.975 \Rightarrow \mathrm{P}(A\cap B)=0.195$ Probability $=\mathrm{P}(A\cap B')+\mathrm{P}(A'\cap B)$ $=[\mathrm{P}(A)-\mathrm{P}(A\cap B)]+[\mathrm{P}(B)-\mathrm{P}(A\cap B)]$ $=0.65-0.195+0.2-0.195$ $=0.46$ OR Probability $=\mathrm{P}(B')\mathrm{P}(A\mid B')+\mathrm{P}(B)\mathrm{P}(A'\mid B)$ $=(0.8)(0.56785)+(0.2)(1-0.975)$ $=0.46$ OR $\mathrm{P}(A'\cap B')=1-\mathrm{P}(A\cup B)$ $=1-[\mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A\cap B)]$ $=1-(0.65+0.2-0.195)$ $=0.345$ Probability $=1-\mathrm{P}(A\cap B)-\mathrm{P}(A'\cap B')$ $=1-0.195-0.345$ $=0.46$	Very common mistake: $\mathrm{P}(A\cap B')+\mathrm{P}(A'\cap B)$ $=\mathrm{P}(A)\mathrm{P}(B')+\mathrm{P}(A')P(B)$ $=(0.65)(0.8)+(0.35)(0.2)$. Student need to realise that the question did not state that *A* and *B* are independent, and they must not assume this. In this question, indeed, *A* and *B* are NOT independent.
9(iii)	$\mathrm{P}(A'\cap B'\cap C')$ $=1-\mathrm{P}(A\cup B\cup C)$ $=1-\left[\begin{array}{l}\mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)-\mathrm{P}(A\cap B)-\mathrm{P}(B\cap C)\\-\mathrm{P}(A\cap C)+\mathrm{P}(A\cap B\cap C)\end{array}\right]$ $=1-[0.65+0.2+0.5-0.195-(0.2)(0.5)-(0.65)(0.5)+0.098]$ $=0.172$	A common mistake: $\mathrm{P}(A'\cap B'\cap C')$ $=\mathrm{P}(A'\cap B')P(C')$, where students assumed that $A'\cap B'$ is independent of *C*'. Some students wrongly interpreted the phrase 'event of either twin being late for practice is independent of the event that C is late for practice' as $A\cup B$ independent of *C*, and wrote $\mathrm{P}((A\cup B)\cap C)=\mathrm{P}(A\cup B)\times\mathrm{P}(C)$.

10	Let X be the random variable for the time taken to place an order and Y be the random variable for the time taken to prepare and serve an order. $X \sim \mathrm{N}(60, \sigma^2), Y \sim \mathrm{N}(300, 50^2)$ Given: $\mathrm{P}(X < 40) = 0.16$ $\Rightarrow \mathrm{P}\left(Z < \frac{40-60}{\sigma}\right) = 0.16$ $\Rightarrow \frac{-20}{\sigma} = -0.9944578907$ $\Rightarrow \sigma = 20.11146$ $\Rightarrow \sigma = 20.11$(to 2 d.p.) (shown)	This is generally well-done. However, no credit is given if student used the value of 20.11 and compared the probabilities of $\mathrm{P}(X < 40)$ for $\sigma = 20.10,\ 20.11$ and 20.12. For a question that requires the student to show the result, student must not use the value to be shown in the working. There are also some students who wrote the standardized value wrongly as $\frac{60-40}{\sigma}$, instead of $\frac{40-60}{\sigma}$.
10(i)	 Since normal curve is symmetrical about mean 60, $\mathrm{P}(49 < X < 55) < \mathrm{P}(57 < X < 63)$. So $A > B$.	Students need to read the question carefully. Many students wrongly interpreted A and B as events rather than probabilities. Hence the following mistakes: 1. Drawing a venn diagram showing A and B as 2 sets 2. Writing $\mathrm{P}(A) > \mathrm{P}(B)$. Some drew 2 different normal curves, without realizing that it's the same random variable X.
10 (ii)	$Y \sim \mathrm{N}(300, 50^2)$ $\mathrm{P}(Y > k) = 0.001$ From GC, $k = 454.5116154 = 455$ s (to 3 s.f.) 	Not well-done despite it being an easy question. Common mistakes: 1. Very careless in interpreting the probability of 0.1%, often writing it as 0.1, or 0.01. 2. Some considered distribution of X, or $X+Y$ instead. Many did additional step of standardizing Y. This is not necessary since the mean and variance of Y are given.

<table>
<tr>
<td>10 (iii)</td>
<td>$\bar{Y} = \dfrac{Y_1 + Y_2 + ... + Y_{10}}{10} \sim \mathrm{N}(300, \dfrac{50^2}{10})$
$\mathrm{P}(\bar{Y} < 270) = 0.028897188$
$= 0.0289$ (to 3 s.f.)</td>
<td>This is generally well-done. Some, however, wrongly quoted the use of CLT.
Common mistakes:
1. Considering distribution of $\dfrac{X_1 + ... + X_{10} + Y_1 + ... + Y_{10}}{10}$
2. Writing variance of $\bar{Y}$ as $10(50^2)$.
3. Not converting 4.5 mins to 270s</td>
</tr>
<tr>
<td>10 (iv)</td>
<td>Let T be random variable for time taken to take the order and prepare the order of 1 customer.
$T = 0.95X + 0.9Y$
$T \sim \mathrm{N}(0.95(60) + 0.9(300), 0.95^2(20.11146^2) + 0.9^2(50^2))$
$T \sim \mathrm{N}(327, 2390.01822)$

Time taken to serve n customers is $T_1 + T_2 + ... + T_n$
$T_1 + T_2 + ... + T_n \sim \mathrm{N}(327n, 2390.01822n)$

Given: $\mathrm{P}(T_1 + ... + T_n < 3600) \geq 0.8$
$\mathrm{P}(Z < \dfrac{3600 - 327n}{\sqrt{2390.01822n}}) \geq 0.8$
$\dfrac{3600 - 327n}{\sqrt{2390.01822n}} \geq 0.84162$
From GC, let $Y_1 = \dfrac{3600 - 327n}{\sqrt{2390.01822n}}$
<table>
<tr><th>n</th><th>Y_1</th></tr>
<tr><td>9</td><td>$4.4796 > 0.84162$</td></tr>
<tr><td>10</td><td>$2.1346 > 0.84162$</td></tr>
<tr><td>11</td><td>$0.0185 < 0.84162$</td></tr>
</table>

Largest $n = 10$

Assume that the time taken to place an order is independent of the time taken to prepare and serve an order for a customer
OR
Assume that the time taken to place an order for a customer is independent of the time taken to place an order for another customer.</td>
<td>Common mistakes:
1. Writing Var(T) as $20.11146^2 + 50^2$ or $(0.95)(20.11146^2) + (0.9)(50^2)$
2. Interpreting the time taken for n customers as nT instead of $T_1 + T_2 + ... + T_n$, and hence writing Var(nT) as $2390.01822n^2$

Students should realise that for normal distributions, independence of random variables is a necessary assumption.
There are students who simply assume that the new improved time for serving customers is normally distributed, and even quote use of CLT.</td>
</tr>
<tr>
<td>11</td>
<td>The probability of a randomly chosen wine glass being chipped is constant.
Selections of chipped wine glasses are independent of each other.</td>
<td>Keywords for <u>assumptions</u>:
• probability … constant
• Selections/events … independent
probability … independent is WRONG!
Note the difference between conditions & assumptions.</td>
</tr>
</table>

Let X be the number of chipped wine glasses in a batch of n glasses.

$X \sim \mathrm{B}(n, p)$

$$\begin{aligned} A &= \mathrm{P}(X \le 1) \\ &= \mathrm{P}(X = 0) + \mathrm{P}(X = 1) \\ &= \binom{n}{0} p^0 (1-p)^n + \binom{n}{1} p^1 (1-p)^{n-1} \\ &= (1-p)^n + np(1-p)^{n-1} \\ &= (1-p)^{n-1}(1-p+np) \\ &= (1-p)^{n-1}\left[1+(n-1)p\right] \text{ (shown)} \end{aligned}$$

Generally well done except –

- Some students mistook A to be $\mathrm{P}(X > 1)$.
- Some only had one of $\mathrm{P}(X = 0)$ or $\mathrm{P}(X = 1)$.
- Some had missing 2nd last step of factoring out $(1 - p)^{n-1}(1 - p + np)$

11 (a)

Given $p = 0.02$, $X \sim \mathrm{B}(n, 0.02)$

$A = \mathrm{P}(\text{batch of glasses will not be rejected}) = \mathrm{P}(X \le 1) \ge 0.9$

From GC,

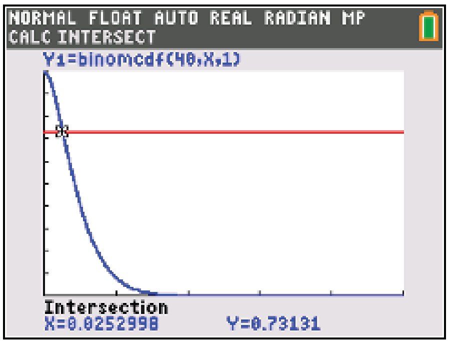
n	$A = P(X \le 1)$
25	$0.9114 > 0.9$
26	$0.9052 > 0.9$
27	$0.8989 < 0.9$

Thus largest $n =$ 26

Generally well done except that some students wasted time in writing out a few steps of working before using GC, instead of comparing with 0.9 directly. Another common mistake is $n = 27$ because of wrong inequality or using equation instead.

11 (b)

Given that $n = 40$, $A = \mathrm{P}(X \le 1) = 0.73131$ where $X \sim \mathrm{B}(40, p)$



From GC, $p = 0.02530$ (5 d.p.)

Common mistake – ignoring '5 decimal places' in the question. Answer such as 0.025300 was common.

11 (b) (i)

$X \sim \mathrm{B}(40, 0.02530)$

$\mu = \mathrm{E}(X) = 40 \times 0.02530 = 1.012$

$\sigma^2 = \mathrm{Var}(X) = 40 \times 0.02530 \times (1 - 0.02530) = 0.9863964$

$$\begin{aligned} &\mathrm{P}(\mu - \sigma < X < \mu + \sigma) \\ &= \mathrm{P}(0.018805 < X < 2.005175) \\ &= \mathrm{P}(X = 1) + \mathrm{P}(X = 2) \\ &= 0.56107 \\ &= 0.561 \text{ (to 3 s.f.)} \end{aligned}$$

Confusion in the distribution of X is common:
$X \sim \mathrm{B}(40, 0.02530)$ at the start becomes
$X \sim \mathrm{N}(1.1012, 0.9863964)$ a few steps later!

A few students took the values of $n = 26$ and $p = 0.02$ from part (a) instead of using the current values stated in part (b) with $n = 40$ and p found in (b) (i)

11 (b) (ii)

Let R be number of rejected batches out of 20×52 batches in a year.

$R \sim \mathrm{B}(20 \times 52, 1 - 0.73131)$

There were a few scripts with no mention of any distributions!

	$R \sim B(1040, 0.26869)$ P(total compensation > \$30000) = $P(R > 300)$ $= 1 - P(R \le 300) = 0.0711$ <u>OR</u> Let C be number of rejected batches out of 20 batches in a week. $C \sim B(20, 1-0.73131)$ $C \sim B(20, 0.26869)$ $E(C) = 20 \times 0.26869 = 5.3738$ $Var(C) = 20 \times 0.26869 \times 0.73131 = 3.929913678$ Since n = 52 is large, by Central Limit Theorem, $C_1 + C_2 + ... + C_{52} \sim N(279.4376, 204.3555113)$ approximately Let total compensation be W. $W = 100(C_1 + ... + C_{52})$ $E(W) = 100(279.4376) = 27943.76$ $Var(W) = 100^2(204.3555113) = 2043555.113$ $W \sim N(27943.76, 2043555.113)$ $P(W > 30000) = 0.075160$ $= 0.0752$ (to 3 s.f.)	Central Limit Theorem was wrongly used for 20 batches instead of 52 weeks. Students should note the random variables involved with their respective n and p: • $X \sim B(40, 0.02530)$ for the number of chipped wine glass • $C \sim B(20, 1 - 0.73131)$ for the number of rejected batches