What does artificial intelligence mean for automatic assessment of mathematics with STACK and Moodle?

Chris Sangwin

School of Mathematics
University of Edinburgh

Feb 2026

Thank you!

Introduction

- What is STACK?
- How do we use STACK?
- Where are we now?
- Technology and education
- Future tools
- Future goals

Key educational message:

We want tests worth teaching to.

Hugh Burkhardt, University of Nottingham

What is STACK?

STACK is a "question type" for mathematics.

- Generate random questions.
- Answers are mathematical.
- Checks the mathematical properties.
- Formative, summative and evaluative outcomes.
  (Text to help, mark/score, data.)

STACK uses computer algebra.

Demonstration

More information about STACK

`https://stack-assessment.org/`

Sorry STACK is difficult to install !

Sorry STACK questions are difficult to write !

Not only mathematics: drag and drop

Prove the following theorem by dragging sentences and arranging them in the correct order.

Theorem: If $3 \cdot 2^{172} + 1$ is a perfect square, then $3 \cdot 2^{172} + 173$ is not a perfect square.

Construct your solution here:

- Assume that $3 \cdot 2^{172} + 1$ is a perfect square.
- Since $3 \cdot 2^{172} + 1 > 2^{172} = (2^{86})^2 > 172^2$, we have $k > 172$.
- There is a positive integer k such that $3 \cdot 2^{172} + 1 = k^2$.
- Also, $3 \cdot 2^{172} + 173 = (3 \cdot 2^{172} + 1) + 172 < k^2 + k$. Further, $k^2 + k < (k+1)^2$.

Drag from here:

- Since $k^2 < 3 \cdot 2^{172} + 173 < (k+1)^2$ it is strictly between two successive squares k^2 and $(k+1)^2$, it cannot be a perfect square.
- We have $k^2 = 3 \cdot 2^{172} + 1 < 3 \cdot 2^{172} + 173$.

Check

Why did I build STACK?

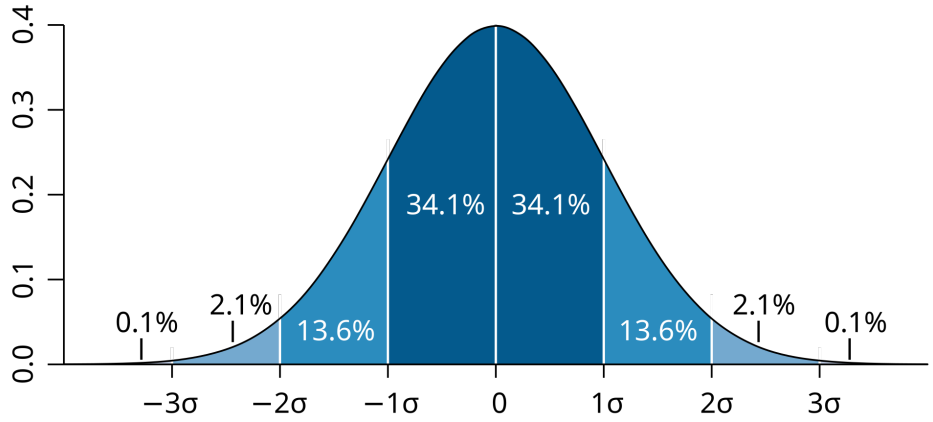
Assessment is the most important part of education.

How do we use STACK?

1. Informal practice
2. Weekly tests
3. Exams
4. Putting the book inside the quiz.
5. (Outside Moodle)

Bloom 1984: learning for mastery

Compare personal tutor with traditional classroom teaching.
Tutor is two standard deviations better on test scores.



M. W. Toews (2005) CC-BY

Learning for mastery

- Students are regularly tested
- Students must score 90%
  (or do more work!)

Now practical with online assessment.

Fundamentals of Algebra and Calculus

- 20 credit year 1 course (20/120=1/6)
- University of Edinburgh.
- 2018–19. Approx 100 students.
- 2019–20. Approx 220 students.
- 2020–21. Approx 220 students.
- ...

  Putting the book inside the online quiz

Used established educational ideas.

- Faded worked examples
- Scaffolding
- Interleaving of topics
- Recall/memory

Assessments

High standards required on weekly online tests
Mastery = 80%+
Distinction = 95%+

A 10 week course.
Final grade:

# Mastered (80%+)	# Distinctions (95%+)	Mark	Grade
< 7	-	0	F
7	-	45%	D
8	2 or 3	55%	C
9	4 or 5	65%	B
10	6 or 7	75%	A1
10	8 or 9	85%	A2
10	10	100%	A3

Demonstrating FAC

`https://stack-demo.maths.ed.ac.uk/demo`

FAC Evaluation

Use of a “diagnostic test”.
Diagnostic test was well established, with good information.

	Pre-test	Post-test	Gain	
Non-FAC	76.1	78.1	2.0	(Control)
FAC	62.1	77.4	+15.3	(Treatment)

Outside Moodle

1. API: exam systems
2. Open Textbooks (Kenya)
3. Sharing questions inside moodle with GIT (on your PC)
   Gitsync plugin
4. (2026-2028: New question bank project)

AI and STACK!

Artificial intelligence is already helping with:

1. Translating questions
2. Authoring questions
3. Mathematical input (writing)
4. Suggesting questions for teachers

Artificial intelligence might/will help with:

1. Feedback
2. Assessing students' work
3. Creating a 'model' of a student
4. Suggesting the next question

Combine STACK and AI

- AI assessment is not clear.
- STACK has clear assessment algorithms.

Combine them: AI $\leftrightarrow$ STACK assessment $\leftrightarrow$ student

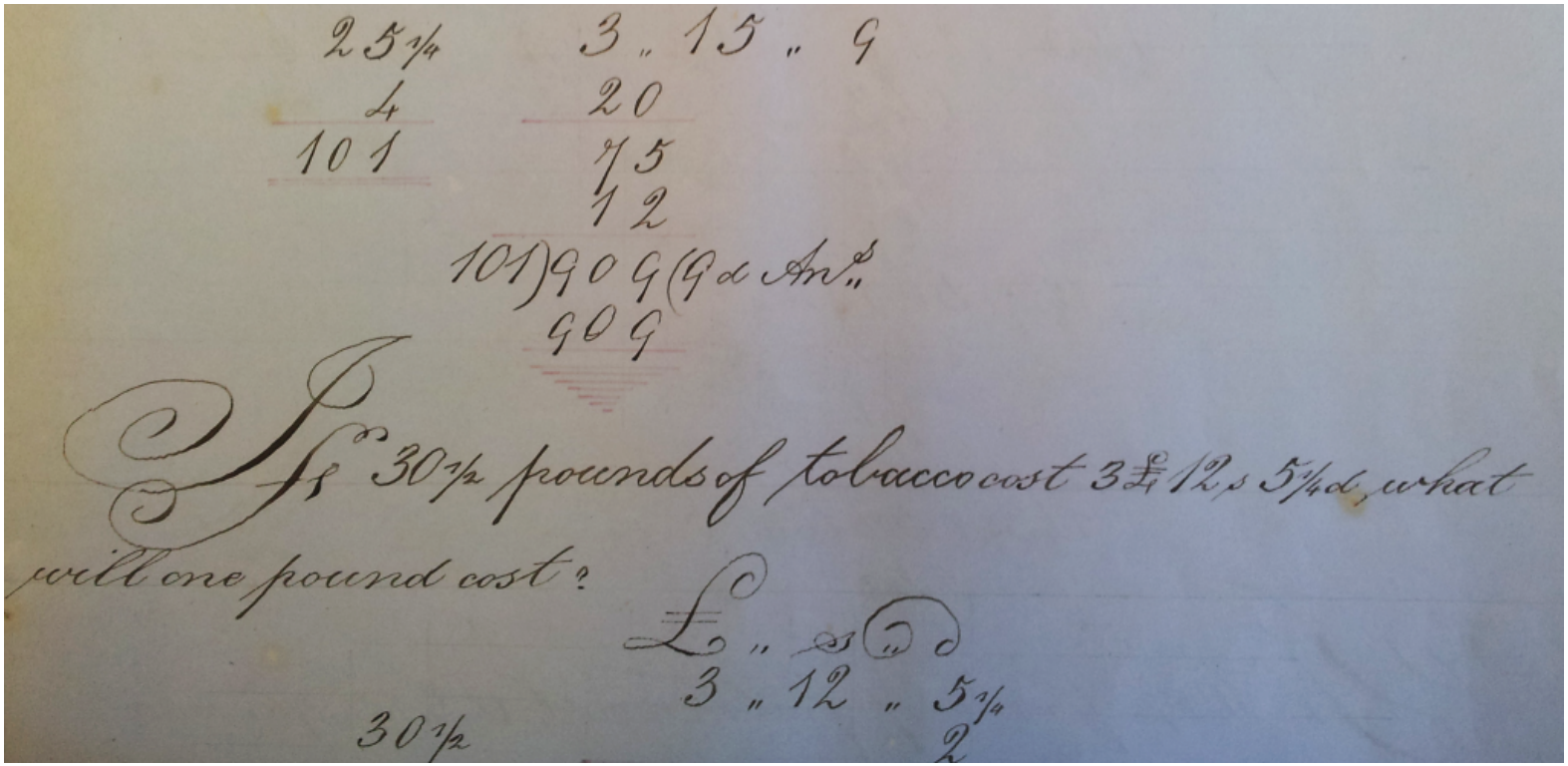
Rules for introducing technology

Hannish's "4 stages"

1. Ignore
2. Forbid
3. Accept (unhappily)
4. Require (expect students to use)

Forbid!

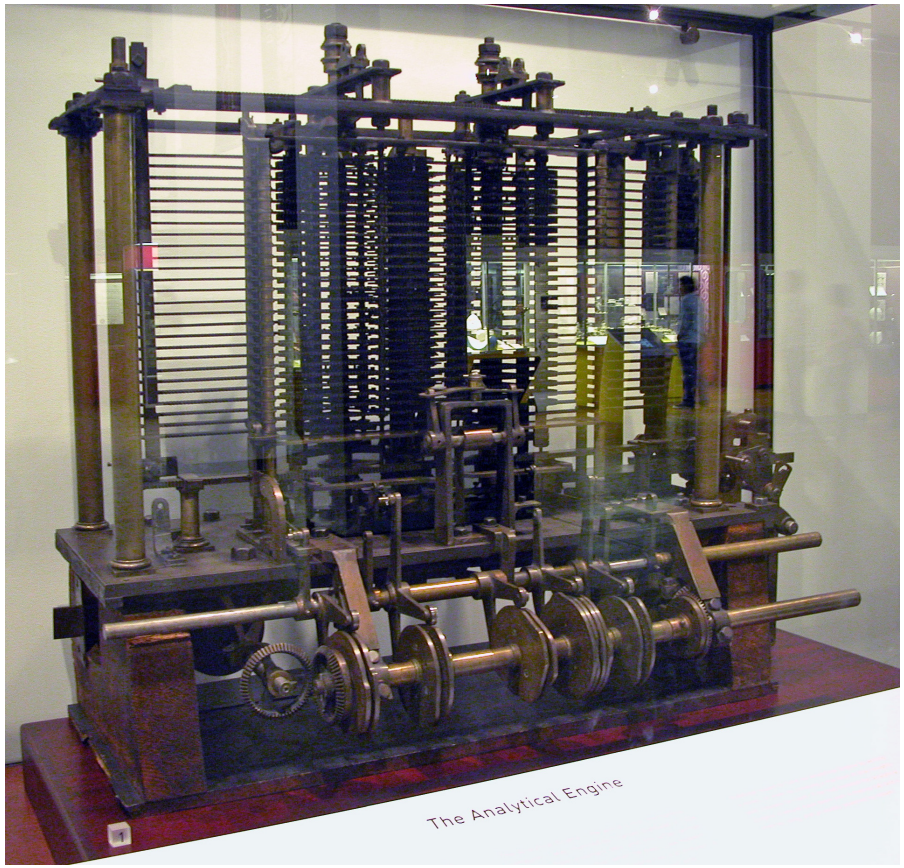
What can the students do themselves?



Joseph Phillip's book 1858 (age 10)

What is he learning?
Why is he learning this?
Do we still need to learn this?

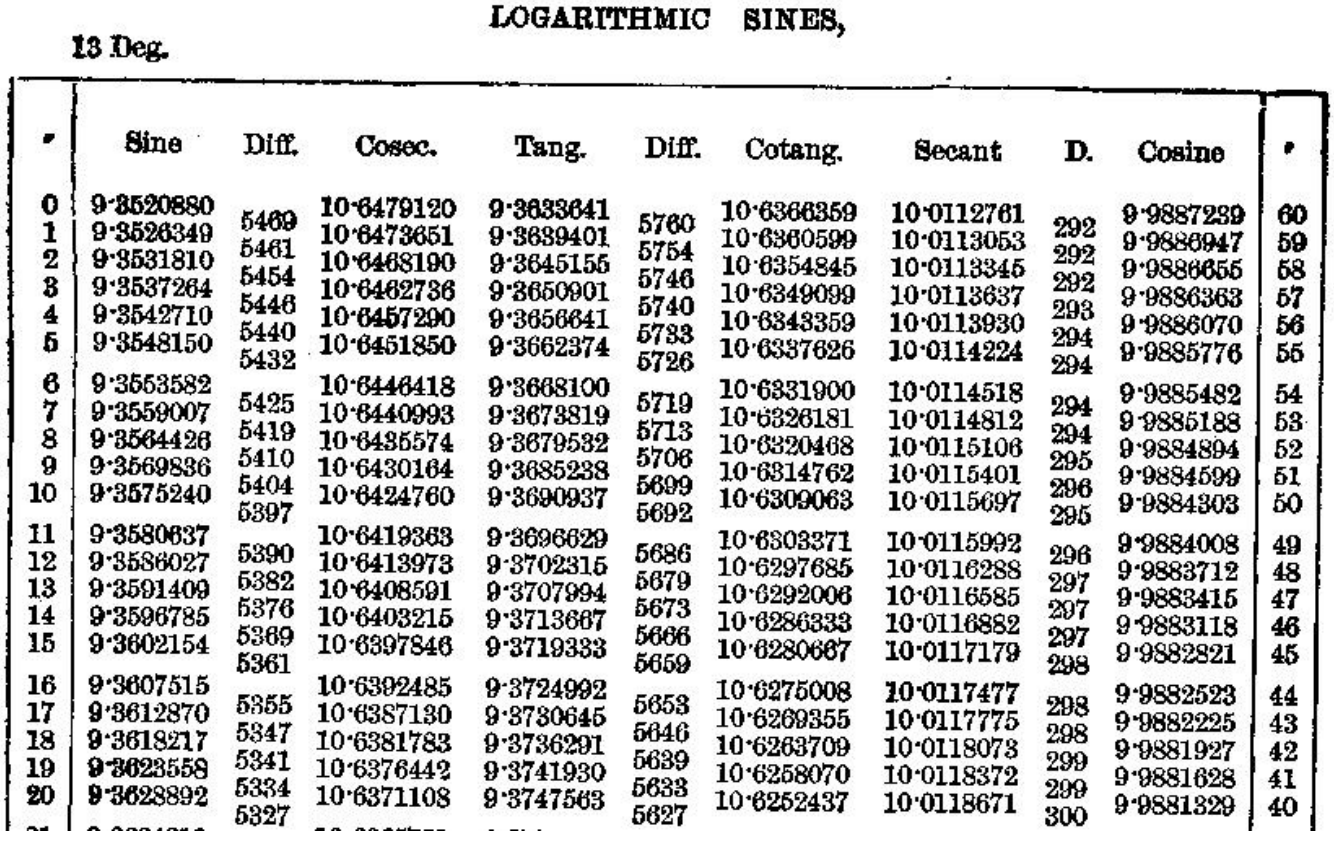
Charles Babbage



This is the *Analytical Engine*. $\approx$ 1837 $\rightarrow$
This is the first mechanical computer.

Technology looks backwards

Babbage wanted to *print log tables!*



LOGARITHMIC SINES,

13 Deg.

′	Sine	Diff.	Cosec.	Tang.	Diff.	Cotang.	Secant	D.	Cosine	′
0	9·3520880	5469	10·6479120	9·3633641	5760	10·6366359	10·0112761	292	9·9887239	60
1	9·3526349	5461	10·6473651	9·3639401	5754	10·6360599	10·0113053	292	9·9886947	59
2	9·3531810	5454	10·6468190	9·3645155	5746	10·6354845	10·0113345	292	9·9886655	58
3	9·3537264	5446	10·6462736	9·3650901	5740	10·6349099	10·0113637	293	9·9886363	57
4	9·3542710	5440	10·6457290	9·3656641	5733	10·6343359	10·0113930	294	9·9886070	56
5	9·3548150	5432	10·6451850	9·3662374	5726	10·6337626	10·0114224	294	9·9885776	55
6	9·3553582	5425	10·6446418	9·3668100	5719	10·6331900	10·0114518	294	9·9885482	54
7	9·3559007	5419	10·6440993	9·3673819	5713	10·6326181	10·0114812	294	9·9885188	53
8	9·3564426	5410	10·6435574	9·3679532	5706	10·6320468	10·0115106	295	9·9884894	52
9	9·3569836	5404	10·6430164	9·3685238	5699	10·6314762	10·0115401	296	9·9884599	51
10	9·3575240	5397	10·6424760	9·3690937	5692	10·6309063	10·0115697	295	9·9884303	50
11	9·3580637	5390	10·6419363	9·3696629	5686	10·6303371	10·0115992	296	9·9884008	49
12	9·3586027	5382	10·6413973	9·3702315	5679	10·6297685	10·0116288	297	9·9883712	48
13	9·3591409	5376	10·6408591	9·3707994	5673	10·6292006	10·0116585	297	9·9883415	47
14	9·3596785	5369	10·6403215	9·3713667	5666	10·6286333	10·0116882	297	9·9883118	46
15	9·3602154	5361	10·6397846	9·3719333	5659	10·6280667	10·0117179	298	9·9882821	45
16	9·3607515	5355	10·6392485	9·3724992	5653	10·6275008	10·0117477	298	9·9882523	44
17	9·3612870	5347	10·6387130	9·3730645	5646	10·6269355	10·0117775	298	9·9882225	43
18	9·3618217	5341	10·6381783	9·3736291	5639	10·6263709	10·0118073	299	9·9881927	42
19	9·3623558	5334	10·6376442	9·3741930	5633	10·6258070	10·0118372	299	9·9881628	41
20	9·3628892	5327	10·6371108	9·3747563	5627	10·6252437	10·0118671	300	9·9881329	40

This normally happens.

E.g. D. Knuth made LaTeX (1989 $\rightarrow$ now....)

Require!

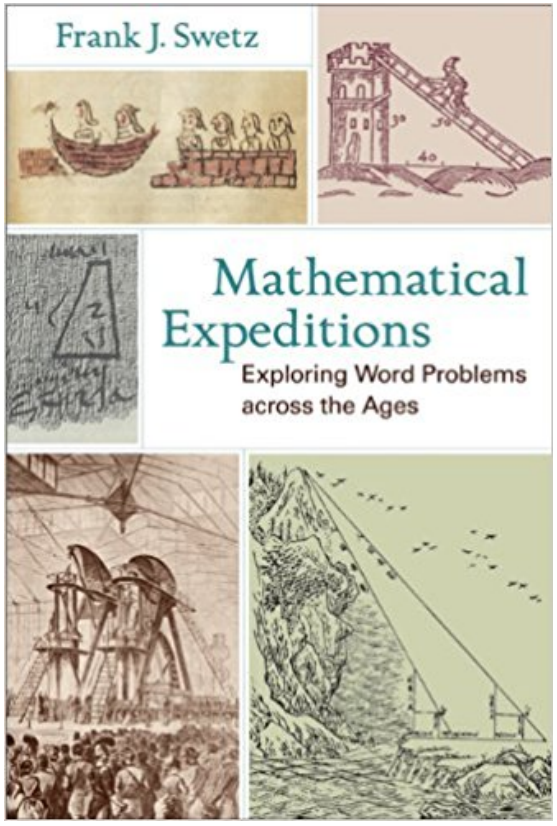
What is left for students to do?

- Applied mathematics:
  - Identifying the problem.
  - Checking answers.
  - Making sense.
- Pure mathematics: culture

Wolf, goat and cabbage problem

A farmer with a wolf, a goat, and a cabbage must cross a river by boat.
The boat can carry only the farmer and a single item.
If left unattended together, the wolf would eat the goat, or the goat would eat the cabbage.
The farmer must help them all cross the river without anything being eaten.

History of problems



- Alcuin of York *Problems to Sharpen the Youth.* (790CE)

Require!

Imagine the AI can solve all the mathematics problems?
What is left?

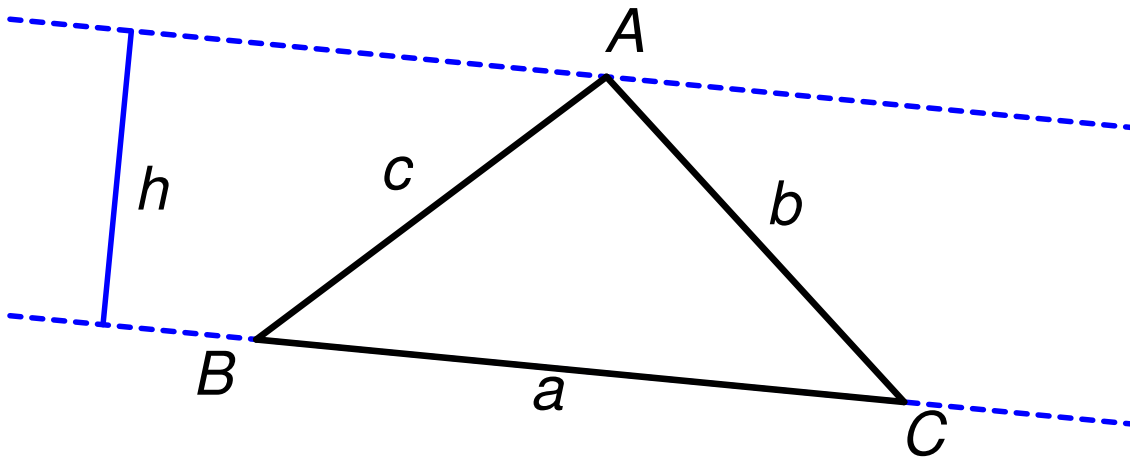
- Magic and surprise
- Patterns $\rightarrow$ mathematical theorems
- Beauty, simplicity, importance
- Being a human participant

We might need to change what we teach to do this....

What is the area of a triangle?

$$\text{area} = \frac{1}{2} \times \text{base} \times \text{height}.$$

$$\Delta = \frac{1}{2}ah$$



The area of a triangle: Baker (1885)

32. $\frac{1}{4}\sqrt{8m_c^2(a^2+b^2-2m_c^2)-(a^2-b^2)^2}$

33. $\frac{1}{2}h_c(\sqrt{a^2-h_c^2}+\sqrt{b^2-h_c^2})$

34. $\frac{1}{2}\beta_c ab\left(\frac{1}{a}+\frac{1}{b}\right)\sqrt{1-\frac{1}{4}\beta_c^2\left(\frac{1}{a}+\frac{1}{b}\right)^2}$

35. $\frac{1}{3}\sqrt{4m_b^2m_c^2-[\frac{9}{4}a^2-(m_b^2+m_c^2)]^2}=\frac{2}{3}\sqrt{m_b^2m_c^2-k_a^4}$

where $2k_a^2=\frac{9}{4}a^2-(m_b^2+m_c^2)$, etc.

36. $\dfrac{h_b^3\sqrt{a^2-h_b^2}-h_c^3\sqrt{a^2-h_c^2}}{2(h_b^2-h_c^2)}$

37. $\frac{1}{6}h_a(\sqrt{4m_b^2-h_a^2}+\sqrt{4m_c^2-h_a^2})$

38. $\frac{1}{2}ah_a$.

39. $\frac{1}{2}\sqrt{abh_ah_b}=\frac{1}{2}\dfrac{a\pm b}{\frac{1}{h_a}\pm\frac{1}{h_b}}$

71. $$\beta_a\beta_b\beta_c \frac{-\beta_a \sin \tfrac{1}{2}A + \beta_b \sin \tfrac{1}{2}B + \beta_c \sin \tfrac{1}{2}C}{-\beta_b\beta_c \cos \tfrac{1}{2}A + \beta_c\beta_a \cos \tfrac{1}{2}B + \beta_a\beta_b \cos \tfrac{1}{2}C}$$

72. $$\tfrac{1}{2}a \sin B\,(a \cos B + \sqrt{b^2 - a^2 \sin^2 B}) = \tfrac{1}{2}b \sin A\,(b \cos A + \sqrt{a^2 - b^2 \sin^2 A})$$

GROUP III. PART I.

73. $$sr = (s - a)\, r_a$$

74. $$\tfrac{1}{3}r\,(\sqrt{2m_a^2 + 2m_b^2 - m_c^2} + \sqrt{2m_b^2 + 2m_c^2 - m_a^2} + \sqrt{2m_c^2 + 2m_a^2 - m_b^2})$$
$$= \tfrac{1}{3}r_a\,(\sqrt{2m_a^2 + 2m_b^2 - m_c^2} - \sqrt{2m_b^2 + 2m_c^2 - m_a^2} + \sqrt{2m_c^2 + 2m_a^2 - m_b^2})$$

75. $$\sqrt{\frac{r r_b r_c}{-\frac{1}{h_a} + \frac{1}{h_b} + \frac{1}{h_c}}} = \sqrt{\frac{r_a r_b r_c}{\frac{1}{h_a} + \frac{1}{h_b} + \frac{1}{h_c}}}$$

76. $$\frac{R}{2s}(h_a h_b + h_b h_c + h_c h_a) = \frac{R}{2(s - a)}(h_a h_b - h_b h_c + h_c h_a)$$

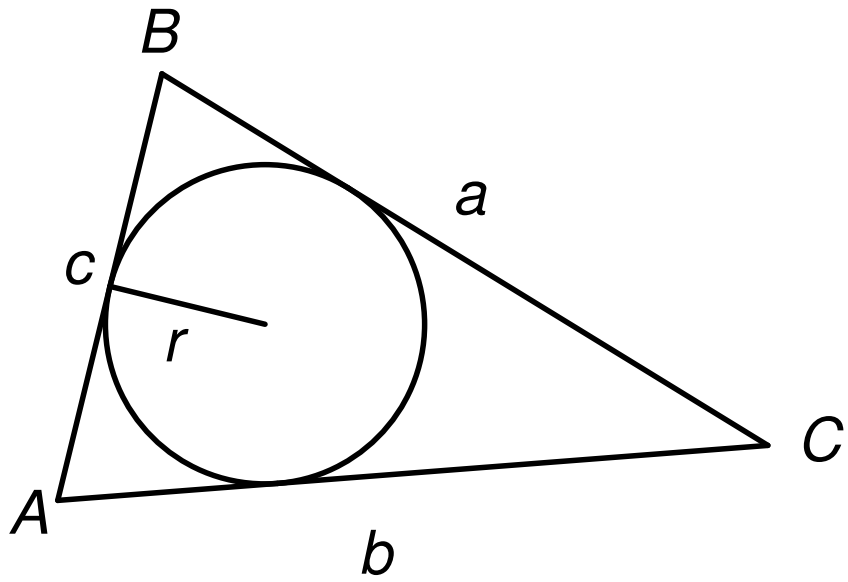
77. $$\sqrt{\tfrac{1}{2}Rr\,(h_a h_b + h_b h_c + h_c h_a)} = \sqrt{\tfrac{1}{2}Rr_a\,(h_a h_b - h_b h_c + h_c h_a)}$$

$\Delta = sr$

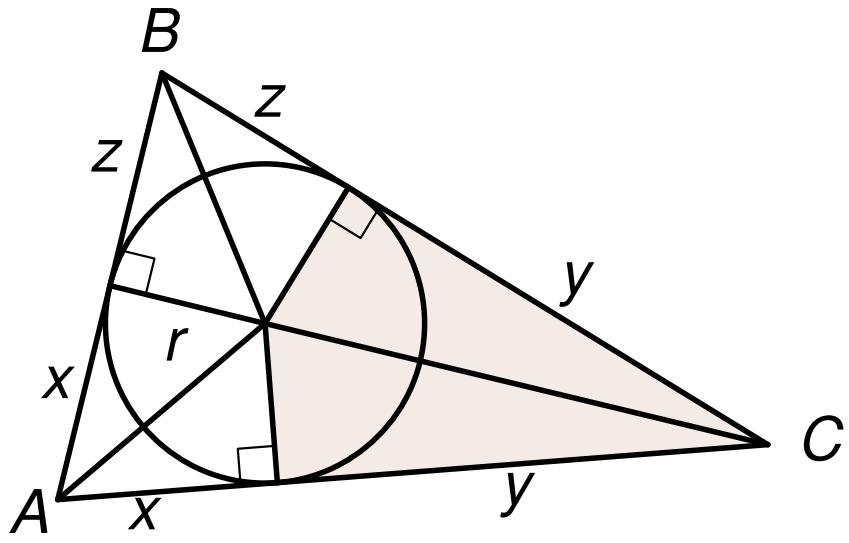
$s =$ half the length of the circumference.

$$s = \frac{a+b+c}{2}$$

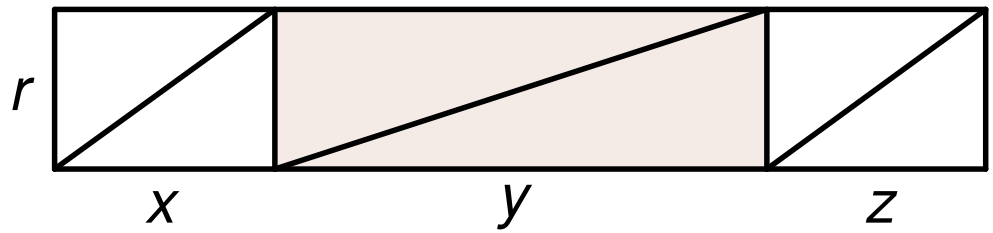
$r =$ radius of the “in-circle”



$\Delta = sr$



$$s = x + y + z$$



$$\Delta = r(x + y + z) = sr.$$

A sequence of problems....

1. Do the same thing with a square: $\Delta = sr$
2. Do the same thing with a hexagon: $\Delta = sr$
3. $\ldots$
4. For a circle:
   - $s = \pi r$, $r = r$.
   - $\Delta = \pi r^2 = sr$
5. For what shapes is $\Delta = sr$?

What does AI mean for STACK and Moodle?

- Sequences of interesting problems worth solving
- Magic and surprise
- Patterns $\rightarrow$ mathematical theorems
- Beauty, simplicity, importance
- Being a human participant
  Humans engaging in mathematical culture

Conclusion

What does artificial intelligence mean for automatic assessment of mathematics with STACK and Moodle?

- We can create better tools (with AI).
- We also need better tests!

  We want tests worth teaching to.