Using online assessment for mathematical proof. Current and future capabilities.

Chris Sangwin

School of Mathematics
University of Edinburgh

November 2021

Thank you for your kind invitation

Outline

1. What is the state of the art in online assessment (STACK)?
2. I will discuss assessment of proof in general.
3. How can we assess students' proofs online today?

What is STACK?

STACK is a "question type" for mathematics.

- STACK generates random questions.
- Students' answers contain mathematical content.
- STACK establishes mathematical properties of students' answers with computer algebra (CAS, Maxima).
- STACK generates formative, summative and evaluative outcomes, (i.e. feedback, score).

Why did I build STACK?

Assessment is the cornerstone of effective education.

- We need assessment worth teaching to.
- I believe universities (we) need to take responsibility for our important tools/software.

System demo

Demonstration of the software.

School exams in STACK?

(Nadine Köcher & Chris Sangwin, 2014)

International Baccalaureate examinations in STACK?

	# marks	
(i) Awarded by STACK (2014) *exactly*	112	18%
(ii) Final answers and implied method marks	227	37%
(iii) Reasoning by equivalence	218	36%
Total of max of (ii) and (iii) per question	376	61%

The most important single form of reasoning in school mathematics is reasoning by equivalence.

Reasoning by equivalence

Work line by line. Lines next to each other are “equivalent".

$$
\begin{aligned}
&\log_3(x+17)-2 &&= \log_3(2x) \quad (x>0, x>-17)\\
\Leftrightarrow &\log_3(x+17)-\log_3(2x) &&= 2\\
\Leftrightarrow &\log_3\left(\frac{x+17}{2x}\right) &&= 2\\
\Leftrightarrow &\frac{x+17}{2x} &&= 3^2 = 9\\
\Leftrightarrow &x+17 &&= 18x\\
\Leftrightarrow &x &&= 1.
\end{aligned}
$$

The above is a *single mathematical object: the argument.*
The above is a single (long) *English sentence.*

Line by line reasoning

Solve $\sqrt{3\,x+4}=2+\sqrt{x+2}$, working line by line. Leave your answer in fully simplified form.

```
sqrt(3*x+4) = 2+sqrt(x+2)
3*x+4 = 4+4*sqrt(x+2)+(x+2)
x-1 = 2*sqrt(x+2)
x^2-2*x+1 = 4*x+8
x^2-6*x-7 = 0
(x-7)*(x+1) = 0
x = 7 or x = -1
x = 7
```

$\sqrt{3\,x+4}=2+\sqrt{x+2}$	$x\in\left[-\frac{4}{3},\,\infty\right)$
$3\,x+4=4+4\,\sqrt{x+2}+(x+2)$	$x\in\left[-2,\,\infty\right)$
$x-1=2\,\sqrt{x+2}$	$x\in\left[-2,\,\infty\right)$
$x^2-2\,x+1=4\,x+8$	
$x^2-6\,x-7=0$	
$(x-7)\,(x+1)=0$	
$x=7 \text{ or } x=-1$	
$x=7$	

The variables found in your answer were: $[x]$

✔ Correct answer, well done.

	$\sqrt{3\,x+4}=2+\sqrt{x+2}$	$x \in \left[-\frac{4}{3},\, \infty\right)$
$\Rightarrow$	$3\,x+4=4+4\,\sqrt{x+2}+(x+2)$	$x \in [-2,\, \infty)$
$\Leftrightarrow$	$x-1=2\,\sqrt{x+2}$	$x \in [-2,\, \infty)$
$\Rightarrow$	$x^2-2\,x+1=4\,x+8$	
$\Leftrightarrow$	$x^2-6\,x-7=0$	
$\Leftrightarrow$	$(x-7)\,(x+1)=0$	
$\Leftrightarrow$	$x=7 \text{ or } x=-1$	
$\Leftarrow$	$x=7$	

Nature of the subject

Polya 1962: *Mathematical Discovery: on understanding, learning and teaching problem solving.*
Polya gives *patterns of thought* for solving problems:

- the pattern of two loci,
- superposition,
- recursion,
- the Cartesian pattern.

Each correct pattern of thought matches a style of proof.

Cartesian pattern

Descartes' *Rules for the Direction of the mind*.

1. Reduce any kind of problem to a mathematical problem.
2. Reduce any mathematical problem to algebra.
3. Reduce any algebra problem to a single equation & solve.

Polya: *"The more you know, the more gaps you can see in this project"*

Solving the equation is only the last step...
Assessment of the whole process is the challenge!

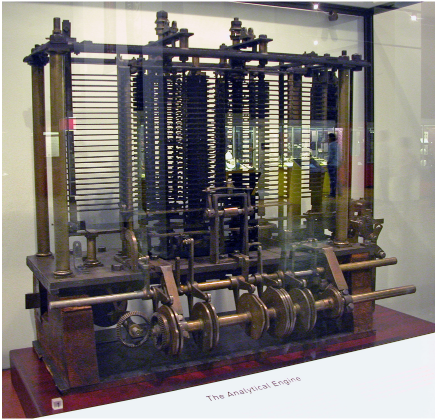
Current State of Freeform-Proof Assessment

Currently there is no really good software for proof-checking.
(Yes, "good" is my personal view!)

Professional automatic reasoning systems. (COQ/LEAN)

But professional mathematicians use LaTeX for papers.

Babbage and the Analytical Engine



This is the *Analytical Engine* invented by Charles Babbage. This is one of the first mechanical computers.

Technology which looks back

Babbage set out 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

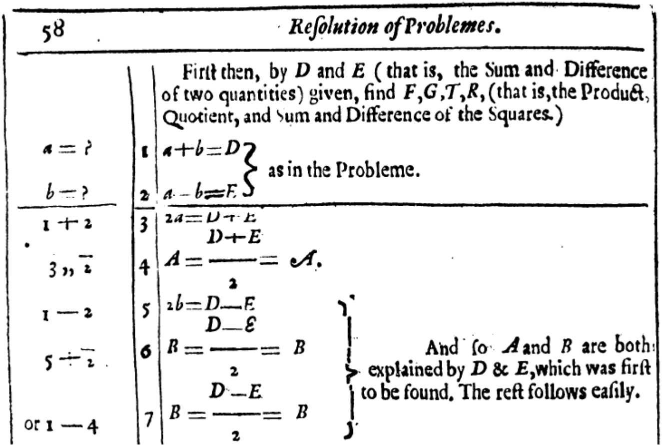
Knuth set out to *replicate movable type!*

Proof: Assessment of a whole argument

Assessment of a complete proof will require a major change in how we write mathematics.

Better interface

In 1668 Pell wrote his proofs using two columns.



58 *Resolution of Problemes.*

First then, by D and E (that is, the Sum and Difference of two quantities) given, find F,G,T,R, (that is, the Product, Quotient, and Sum and Difference of the Squares.)

$a = ?$	1	$a+b=D$ } as in the Probleme.
$b = ?$	2	$a-b=E$
1 + 2	3	$2a=D+E$
3 ,, $\bar{2}$	4	$A=\frac{D+E}{2}=\mathcal{A}.$
1 — 2	5	$2b=D-E$
5 ÷ $\bar{2}$	6	$B=\frac{D-E}{2}=B$
or 1 — 4	7	$B=\frac{D-E}{2}=B$

} And so A and B are both explained by D & E, which was first to be found. The rest follows easily.

Pell (1668) (see Stedall (2002))

Writing in two columns is not popular for school mathematics (UK).

Reasoning by equivalence demo

Replicating standard practice:

Solve $\sqrt{3\,x+4} = 2+\sqrt{x+2}$, working line by line. Leave your answer in fully simplified form.

```
sqrt(3*x+4) = 2+sqrt(x+2)
3*x+4 = 4+4*sqrt(x+2)+(x+2)
x-1 = 2*sqrt(x+2)
x^2-2*x+1 = 4*x+8
x^2-6*x-7 = 0
(x-7)*(x+1) = 0
x = 7 or x = -1
x = 7
```

$\sqrt{3\,x+4} = 2+\sqrt{x+2}$	$x \in \left[-\frac{4}{3}, \infty\right)$
$3\,x+4 = 4+4\,\sqrt{x+2}+(x+2)$	$x \in [-2, \infty)$
$x-1 = 2\,\sqrt{x+2}$	$x \in [-2, \infty)$
$x^2-2\,x+1 = 4\,x+8$	
$x^2-6\,x-7 = 0$	
$(x-7)\,(x+1) = 0$	
$x = 7 \text{ or } x = -1$	
$x = 7$	

The variables found in your answer were: $[x]$

✔ Correct answer, well done.

$$\begin{array}{lll}
 & \sqrt{3\,x+4}=2+\sqrt{x+2} & x\in\left[-\frac{4}{3},\,\infty\right) \\
\Rightarrow & 3\,x+4=4+4\,\sqrt{x+2}+(x+2) & x\in[-2,\,\infty) \\
\Leftrightarrow & x-1=2\,\sqrt{x+2} & x\in[-2,\,\infty) \\
\Rightarrow & x^2-2\,x+1=4\,x+8 & \\
\Leftrightarrow & x^2-6\,x-7=0 & \\
\Leftrightarrow & (x-7)\,(x+1)=0 & \\
\Leftrightarrow & x=7 \text{ or } x=-1 & \\
\Leftarrow & x=7 &
\end{array}$$

Encounters with proof

The short-term goal is not to assess students' proof.

Assessing components of a proof might better serve students.

Classical ways to reduce the difficulty (cognitive load)

1. Hints.
2. Split complex problem into parts.

We don't do the following very much (in the UK).

3. Fill-in the missing gaps.
4. Faded worked examples.
5. Separated concerns.
6. Reading comprehension.

Example of proof with gaps

Let $P(n)$ be the statement

$\sum_{k=1}^{n} k^2 = \frac{n\cdot(n+1)\cdot(2\cdot n+1)}{6}$

Since $1^2 = 1$ [1]

and $\frac{1\cdot(1+1)\cdot(2\cdot 1+1)}{6} = 1$ [1] ✔

it follows that $P(1)$ is true.

Assume that $P(n)$ is true.

$\sum_{k=1}^{n+1} k^2 = \sum_{k=1}^{n} k^2 + (n+1)^2$ [sum(k^2,k,1,n)+(n+1)^2]

$= \frac{n\cdot(n+1)\cdot(2\cdot n+1)}{6} + (n+1)^2$ [(by the induction hypothesis)] ✔

$= \frac{(n+2)\cdot(2\cdot n+3)\cdot(n+1)}{6}$

$= \frac{(n+1)\cdot(n+1+1)\cdot(2\cdot(n+1)+1)}{6}$ [((n+1)*((n+1)+1)*(2*(n+1)+1))/( ] ✔

Since $P(1)$ and $P(n) \Rightarrow P(n+1)$ it follows that $P(n)$ is true for all $n \in \mathbb{N}$ by the principal of mathematical induction.

Example of proof with gaps

Complete the following proof.

Theorem: Let $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ be three linearly independent vectors. Then

$$\mathrm{span}\{\mathbf{x},\mathbf{y},\mathbf{z}\} = \mathrm{span}\{\mathbf{x}+\mathbf{y},\mathbf{y}+\mathbf{z},\mathbf{z}+\mathbf{x}\}.$$

Proof.

Assume $\mathbf{v}$ [∈ ⬍] $W := \mathrm{span}\{\mathbf{x}+\mathbf{y},\mathbf{y}+\mathbf{z},\mathbf{z}+\mathbf{x}\}$ then there exist $a, b, c \in \mathbb{R}$ such that

$\mathbf{v} = a(\mathbf{x}+\mathbf{y}) + b(\mathbf{y}+\mathbf{z}) + c(\mathbf{z}+\mathbf{x})$

$=$ [c+a] $\mathbf{x}+$ [b+a] $\mathbf{y}+$ [c+b] $\mathbf{z}$

so that if $\mathbf{v}$ [∈ ⬍] W then $\mathbf{v} \in U$, i.e. W [⊆ ⬍] U.

Let $U := \mathrm{span}\{\mathbf{x},\mathbf{y},\mathbf{z}\}$. Assume $\mathbf{v}$ [∈ ⬍] [∈ / ∉ / = / ⊆ / ⊇ / ⊈ / ⊉ / ≤ / ≥] n there exist $a, b, c \in \mathbb{R}$ such that $\mathbf{v} = a\mathbf{x} + b\mathbf{y} + c\mathbf{z}$

$=$ [(-c+b+a)/] $(\mathbf{x}+\mathbf{y})+$ [(c+b-a)/2] $(\mathbf{y}+\mathbf{z})$ [+a)/2] $(\mathbf{z}+\mathbf{x})$.

That is to say, U [⊆ ⬍] W.

Hence $U = W$ and

$$\mathrm{span}\{\ldots\} = \mathrm{span}\{\mathbf{x}+\mathbf{y},\mathbf{y}+\mathbf{z},\mathbf{z}+\mathbf{x}\}.$$

Faded worked examples

1. A sequence of questions.
2. Students do more with each step.
3. The long-term goal is students become completely independent.

While there is no suggestion we can mark a complete proof online.
I think students can come to class better prepared.

Separation of concerns

There is a lot going on in a typical proof!
For example

1. *Logical status of statements and proof framework.*
2. *Meaning of terms and statements within the proof.*
3. *Justification of claims.*
4. *Summarizing via high-level ideas.*
5. *Identifying the modular structure.*
6. *Transferring the general ideas or methods to another context.*
7. *Illustrating with examples.*

(Mejia-Ramos 2012)

Separate concerns example

Let $P(n)$ be the statement $\sum_{k=1}^{n} k \cdot k! = (n+1)!$

1. Write the statement $P(n+1)$: sum(?,k,1,?)=?

2. Calculate

$$\sum_{k=1}^{n+1} k \cdot k! - \sum_{k=1}^{n} k \cdot k!$$

writing your answer in simplified form.

3. Calculate

$$(n+2)! - (n+1)!$$

writing your answer in simplified form.

4. Assume $n \in \mathbb{N}$. For which values of n is the implication $P(n) \Rightarrow P(n+1)$ true?

You may answer with an inequality, e.g. `n>?`, a set `{?, ?, ..., ?}`, or with `N` $= \mathbb{N}$ or `none`.

5. For which values of $n \in \mathbb{N}$ is the $P(n)$ true?

You may answer with an inequality, e.g. `n>?`, a set `{?, ?, ..., ?}`, or with `N` $= \mathbb{N}$ or `none`.

Is this trivial for students?

Only 45% of our year 1 students correctly evaluate

$$(n+2)! - (n+1)! = (n+1)(n+1)!$$

The separated concerns example is not trivial for our students.
If students complete the CAA correctly before they write a traditional induction proof they will learn more.

Reading comprehension

Ask students *about* a particular proof.
We found it quite hard to write these questions.

Proof understanding baseline checklist

1. Which formal definitions/notations are relevant to the proof?
2. Describe the overall nested structure of the proof.
3. Hypotheses
    1. Where is each hypothesis used in the proof?
    2. In a general proof, which examples do/do not satisfy the hypotheses? If there is more than one hypothesis, do we have examples which satisfy each logical combination?
4. Is a correct warrant justifying each step in the proof given? If not then provide one.
5. Does the proof make use of previously known theorems or results? If so, what are they and how are they used?
6. Does the proof make use of proof-gadgets? If so, what are they and how are they used?
7. For an if ... then proof, is the converse true or false? Do we have counter-examples?
8. In a general proof, can you follow the proof steps through with a simple specific example, including any proof-gadgets?

Proof understanding baseline checklist

1. Which formal definitions/notations are relevant to the proof?
2. Describe the overall nested structure of the proof.
3. Hypotheses
    1. Where is each hypothesis used in the proof?
    2. In a general proof, which examples do/do not satisfy the hypotheses? If there is more than one hypothesis, do we have examples which satisfy each logical combination?
4. Is a correct warrant justifying each step in the proof given? If not then provide one.
5. Does the proof make use of previously known theorems or results? If so, what are they and how are they used?
6. Does the proof make use of proof-gadgets? If so, what are they and how are they used?
7. For an if ... then proof, is the converse true or false? Do we have counter-examples?
8. In a general proof, can you follow the proof steps through with a simple specific example, including any proof-gadgets?

Nested structure of a proof?

Traditional proof:

Theorem

If $a + b\sqrt{2} = c + d\sqrt{2}$ and $a, b, c, d \in \mathbb{Q}$ then $a = c$ and $b = d$.

Proof.

Suppose (for a contradiction) that $b \neq d$. If $a + b\sqrt{2} = c + d\sqrt{2}$ then, rearranging, we have $(a - c) = (d - b)\sqrt{2}$. Dividing gives $\sqrt{2} = \frac{a-c}{d-b} \in \mathbb{Q}$. But [as previously proved] $\sqrt{2} \notin \mathbb{Q}$. This is a contradiction, so $b = d$. Then setting $b = d$ in $a + b\sqrt{2} = c + d\sqrt{2}$ it follows $a = c$. □

More structured

Proof.

Assume $a + b\sqrt{2} = c + d\sqrt{2}$ and $a, b, c, d \in \mathbb{Q}$. Then

$$
\begin{aligned}
& a + b\sqrt{2} = c + d\sqrt{2} \\
\Leftrightarrow\ & (a - c) = (d - b)\sqrt{2}.
\end{aligned}
$$

1. If $b \neq d$ then $\sqrt{2} = \frac{a-c}{d-b}$. Since $a, b, c, d \in \mathbb{Q}$ it follows $\frac{a-c}{d-b} \in \mathbb{Q}$. But [as previously proved] $\sqrt{2} \notin \mathbb{Q}$. This contradicts the assumption $b \neq d$.
2. If $b = d$ then $(a - c) = 0$, i.e. $a = c$, and the theorem holds.

The only case which holds is $b = d$ and so $a = c$. □

Explicit structure

Equivalence reasoning.
Cases:

- $b \neq d$: *Contradiction.*
- $b = d$: *Direct proof.*

Proof understanding baseline checklist

1. Which formal definitions/notations are relevant to the proof?
2. Describe the overall modular recursive structure of the proof.
3. Hypotheses
    1. Where is each hypothesis used in the proof?
    2. In a general proof, which examples do/do not satisfy the hypotheses? If there is more than one hypothesis, do we have examples which satisfy each logical combination?
4. Is a correct warrant justifying each step in the proof given? If not then provide one.
5. Does the proof make use of previously known theorems or results? If so, what are they and how are they used?
6. Does the proof make use of proof-gadgets? If so, what are they and how are they used?
7. For an if ... then proof, is the converse true or false? Do we have counter-examples?
8. In a general proof, can you follow the proof steps through with a simple specific example, including any proof-gadgets?

Proof gadgets

“a device within a proof, built to establish certain conditions must hold.”
E.g. proof of infinitely many primes

$$N = p_1 p_2 \cdots p_n + 1$$

Proof understanding baseline checklist

1. Which formal definitions/notations are relevant to the proof?
2. Describe the overall modular recursive structure of the proof.
3. Hypotheses
    1. Where is each hypothesis used in the proof?
    2. In a general proof, which examples do/do not satisfy the hypotheses? If there is more than one hypothesis, do we have examples which satisfy each logical combination?
4. Is a correct warrant justifying each step in the proof given? If not then provide one.
5. Does the proof make use of previously known theorems or results? If so, what are they and how are they used?
6. Does the proof make use of proof-gadgets? If so, what are they and how are they used?
7. For an if ... then proof, is the converse true or false? Do we have counter-examples?
8. In a general proof, can you follow the proof steps through with a simple specific example, including any proof-gadgets?

Which examples do/do not satisfy the hypotheses?

Theorem: If (a_n) is a bounded and increasing sequence then $\lim_{n\to\infty} a_n$ exists.

Inc ?	**Bdd ?**	**Con ?**	**Example**
T	T	T	Exemplify theorem: $a_n = 1 - \frac{1}{n}$
T	T	F	Counter example!
T	F	T	Note A.
T	F	F	$a_n = n$
F	T	T	$a_n = 1/n$
F	T	F	$a_n = (-1)^n$
F	F	T	Note A.
F	F	F	$a_n = (-n)^n$

Note A: Bounded is a necessary condition for convergence.

Encounters with proof

Valuable activities associated with proof.

Writing sequences of problems

... is something of an art form.
It is much easier to ask students to “prove this...”!

Conclusion

Computer aided assessment of mathematics: the current state of the art and a look to the future.

- We can largely automate the methods-based parts.
- Increasingly asking *about* proof and reasoning.
- We might better serve students with careful *encounters with proofs*.
- Assessment of free-form proof is some way off, but online submission and human marking does have its place.