Representing Image-Required Mathematics for Intelligent Textbooks

Worcester Polytechnic Institute

Diagram encodings for intelligent textbook services

Ethan Croteau and Neil Heffernan

ecroteau@wpi.edu | nth@wpi.edu

7th Workshop on Intelligent Textbooks | June 28, 2026

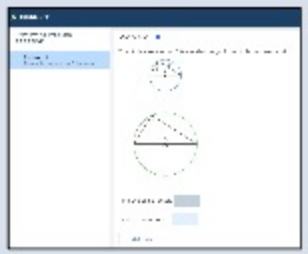
Core claim

Diagram encodings are design choices for intelligent textbook services, not interchangeable file formats.

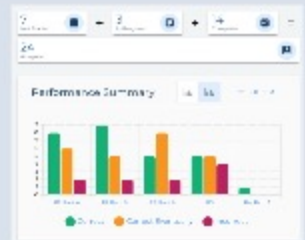
ASSISTments

A digital tool for independent math practice and formative assessment, founded by Neil and Cristina Heffernan in 2003. It is now being scaled by **The ASSISTments Foundation**, a nonprofit advancing math instruction for students in grades 3–8 and their teachers. Used by over 2,300 teachers and their over 124,000 students in 2025-26 SY.

Students practice with immediate feedback.
Hints and explanations help them learn from mistakes in the moment.



Teachers adjust instruction with real-time data.
Reports surface misconceptions by student, problem, skill, and standard.



Researchers use ASSISTments in two ways (etrialstestbed.org)

1. Secondary data analysis

Researchers frequently use released datasets such ASSIST-2009-10 or most recently FoundationalASSIST, among others. find more datasets at: etrialstestbed.org/data-sets

2. Primary data collection

Design and run your own studies. Build content and deliver it to students through ASSISTments using Self-Deploy or Auto-Deploy methods, with built-in randomization support.

WPI

The practical problem

Static images are not enough for the services iTextbooks want to provide.

Tutor or QA system

Needs diagram facts to ground hints, explanations, and refusals.

Accessibility layer

Needs visible structure described without accidentally giving away the solution.

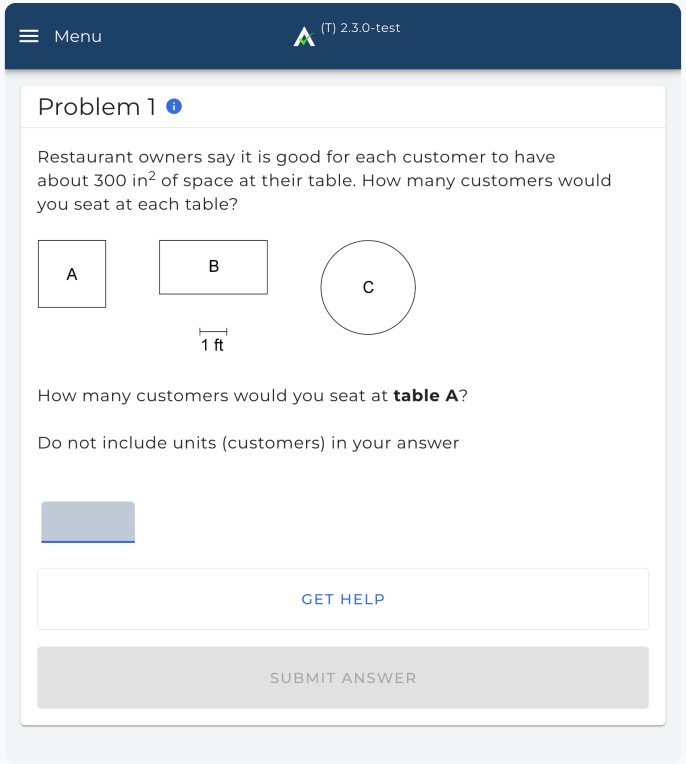
Authoring / validation

Needs inspectable source that can be checked, adapted, and re-rendered.

A learner-facing image can be the right representation for a student and the wrong representation for a machine service.

One textbook item, two kinds of evidence

The written prompt stays the same; the diagram supplies facts needed to solve it.



Problem text

How many customers would you seat at table A?

Diagram evidence

- Table A is square.
- The scale bar converts image length to feet.
- Area per customer is computed from visual size.

Changing the diagram representation changes what mathematical information the system can use.

What makes it image-required?

The written prompt is incomplete without facts supplied by the diagram.

Problem text alone

How many customers would you seat at table A?

This question cannot be answered from the words alone.

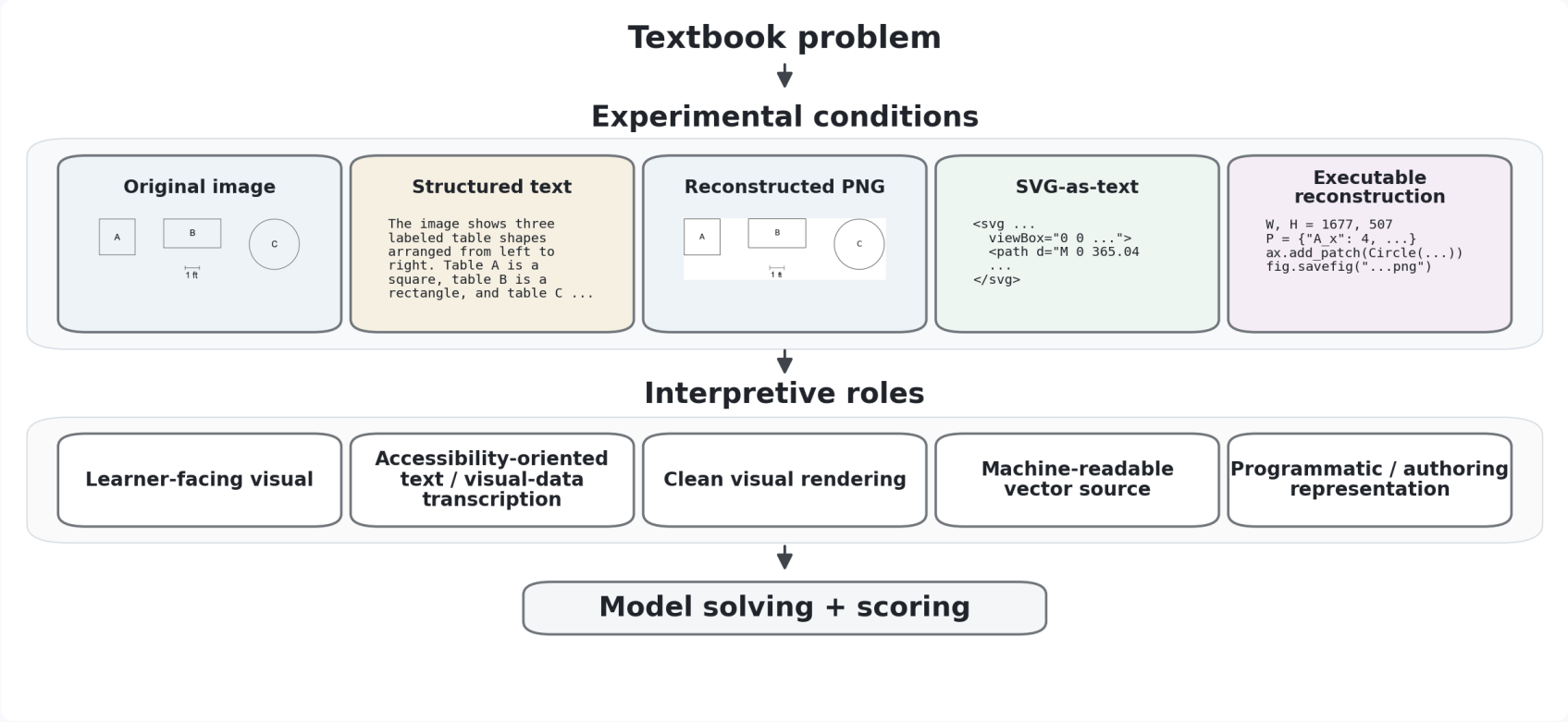
Missing without the diagram

- shape of table A
- scale from image length to feet
- usable area for seating

Image-required means the diagram supplies facts needed to solve the problem.

Five encodings, one problem text

The experiment changes only the diagram representation.



Interpretive point: these are service representations, not claims that every version is visually equivalent.

Three research questions

Accuracy is the measurement task; representation design is the textbook question.

RQ1 How does performance vary across diagram encodings?

RQ2 How does semantic enrichment relate to performance?

RQ3 What do problem-level effects imply for textbook services?

Contributions:

1. A controlled pilot study: same problem text, same scored tasks, but different diagram representations
2. An audit vocabulary: a set of terms for judging diagram representations

Controlled comparison design

Exploratory answer-scored study with the original problem text held fixed.

34
image-required middle-school math problems

5
diagram encodings per problem

2
pinned model snapshots

3
attempts per problem-condition-model cell

34 problems × 5 representations × 2 models × 3 attempts = 1,020 scored runs

- Held fixed: problem text, LLM prompt, and scoring.
- Problem-cluster bootstrap intervals summarize descriptive uncertainty.

Representations were prepared and checked

Each encoding was reviewed for faithfulness, neutrality, and explicit structure.

Interpretive caution

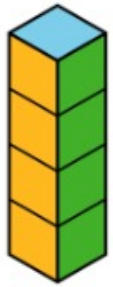
Some encodings may surface measurements or structure a learner would normally extract visually. Good for some services, but it changes the task.

The sample spans six visual demands

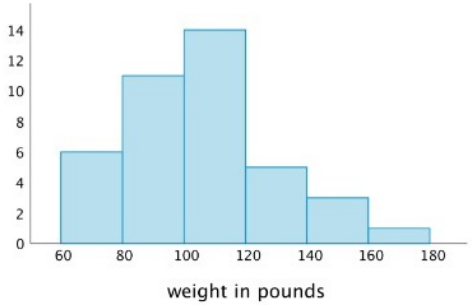
Counts are purposeful sample counts, not prevalence estimates.

4 Measurement and scale

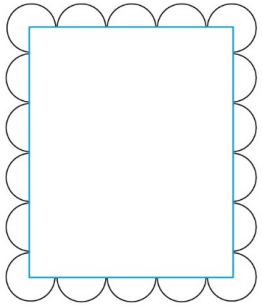
5 Counting and enumeration



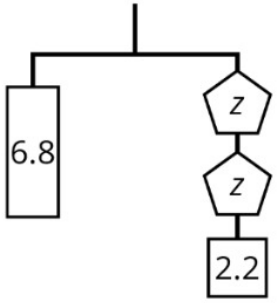
10 Graphs and coordinates



8 Geometric structure



4 Relational diagrams

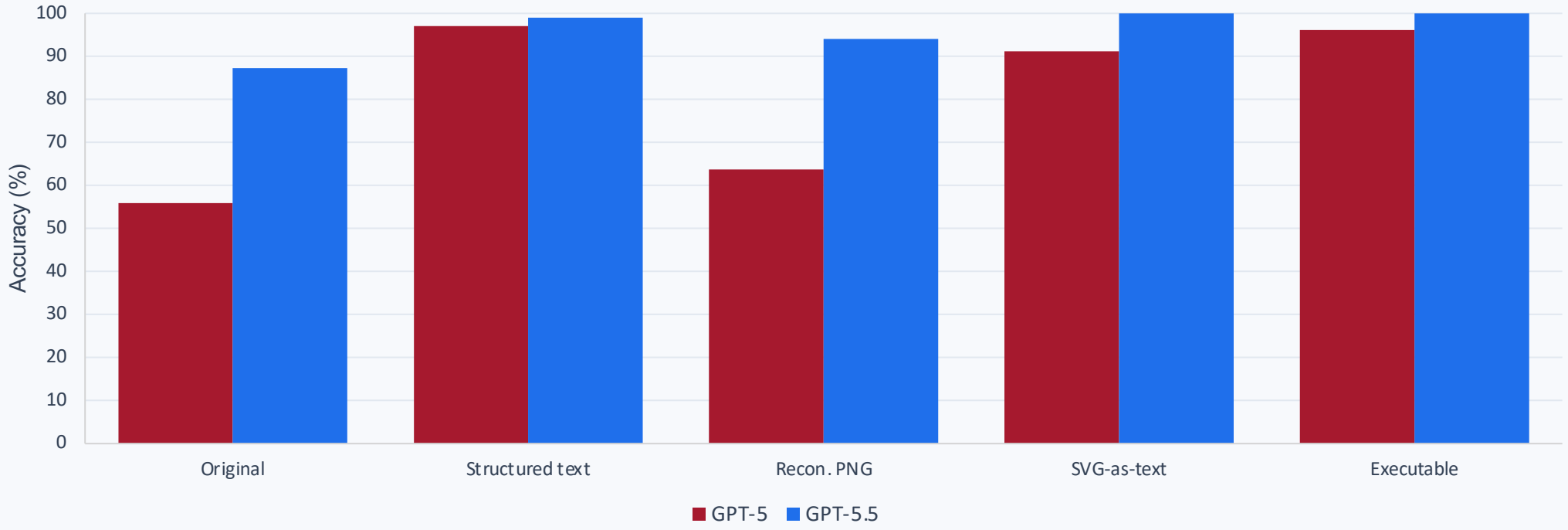


3 Tables and proportions

honey (tbsp)	flour (c)
1	A
8	10
16	B
20	C
h	D

Result 1: encoding choice changes accuracy

Semantically explicit encodings outperform image-only baselines in this pilot.



For GPT-5, accuracy rises from 55.9% with original images to 97.1% with structured text.

Visual cleanup is not the whole story

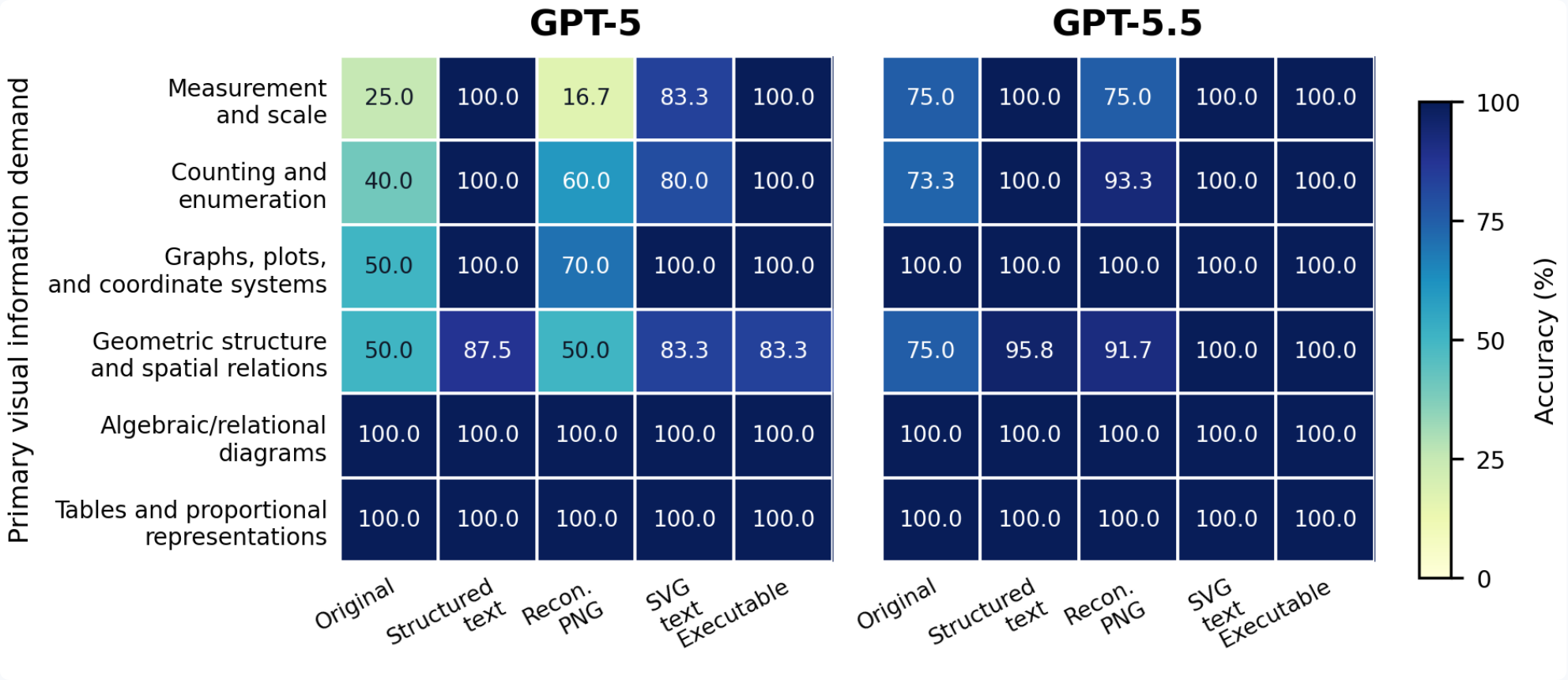
A cleaner PNG helps less than representations that expose task structure.

Encoding	GPT-5 delta	GPT-5.5 delta	Interpretation
Structured text	+41.2 pts	+11.8 pts	task facts explicit
Reconstructed PNG	+7.8 pts	+6.9 pts	cleaner visual
SVG-as-text	+35.3 pts	+12.7 pts	vector structure exposed
Executable	+40.2 pts	+12.7 pts	programmatic structure

The largest gains come from exposing task structure, not just from cleaner images.

Effects also depend on visual demand

The same encoding does not help every kind of diagram equally.



Category macro averages preserve the same ordering, but sparse categories mean this is descriptive.

Result 2: semantic enrichment matters

Medium and high enrichment expose structure a solver would otherwise extract visually.

Important nuance: more enrichment is not automatically better; its appropriateness depends on the service.

Result 3: problem-level effects are heterogeneous

Averages hide robust items, rescued items, and cases where alternatives hurt.

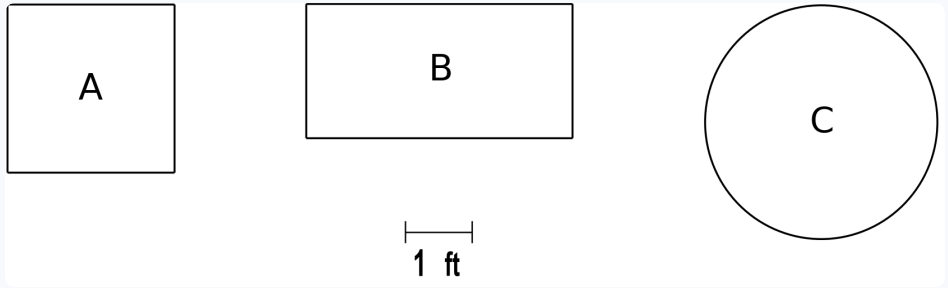
Model	All correct	Rescued by alt.	Improved by alt.	Reduced by alt.	None correct
GPT-5	16	12	5	1	0
GPT-5.5	27	3	2	2	0

Design implication

A textbook should validate representations at the problem level. A condition-level winner can still be wrong for a particular item.

Case A: structure can rescue a measurement task

Table A shows the benefit and the validity risk at the same time.



Model	Orig.	Text	PNG	SVG	Code
GPT-5	1/3	3/3	0/3	1/3	3/3
GPT-5.5	0/3	3/3	3/3	3/3	3/3

Structured text says:

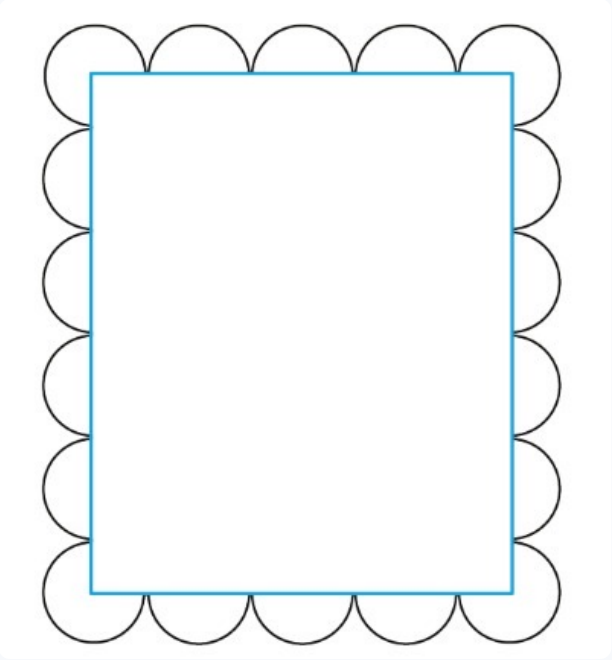
"square A has a side length of about 2.5 ft"

Benefit **Externalizes the measurement.**

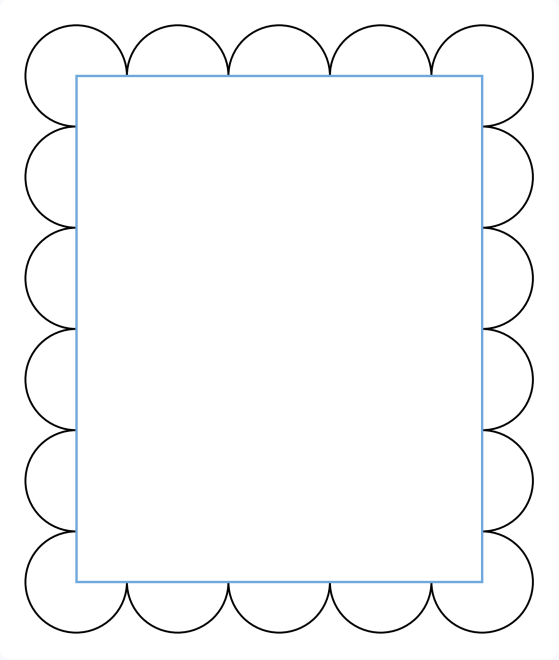
Risk **Not equivalent to visual measurement.**

Case B: code and SVG can help geometry

A scalloped boundary problem remains hard from the image alone.



Original image



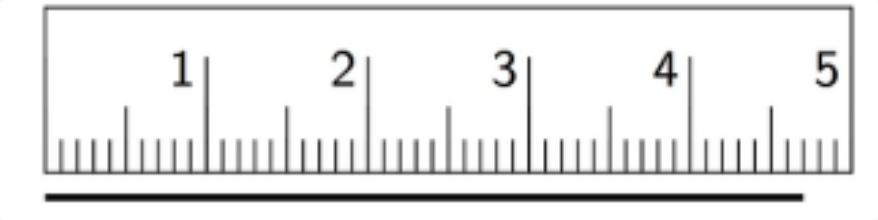
Reconstructed PNG

Condition	GPT-5.5 correct
Original	0/3
Structured text	2/3
Reconstructed PNG	1/3
SVG-as-text	3/3
Executable	3/3

Here, vector/code structure clarifies the hidden construction of the boundary.

Case C: reconstruction can also hurt

Even low-enrichment visual reconstructions need validation.



Original image

Reconstructed PNG

Model	Original	Recon. PNG	Text, SVG, code
GPT-5	1/3	0/3	3/3
GPT-5.5	3/3	1/3	3/3

A cleaner image is not always a better model input.

Implications for intelligent textbooks

Choose diagram encodings by service, then validate them.

Service	Representation need	Likely representation
Learner display	faithful visual form	original or reconstructed PNG
Accessibility / QA	explicit visible facts	structured text
Tutoring	grounded hints without over-disclosure	text + image + guardrails
Authoring / validation	inspectable construction	SVG or executable code
Benchmarking	separate perception from reasoning	multiple encodings

The output is a representation policy, not a winning file format.

Takeaways and Q&A

Three points to leave on screen for discussion.

1 **Visual mathematics content needs layered, validated diagram representations.**

2 **Accuracy gains often reflect semantic enrichment, not just better image understanding.**

3 **Representation choice should be service-specific and validated at the problem level.**

Supplementary materials

https://osf.io/5wbxd/