

TEACHER RESEARCH • AI IN THE CLASSROOM

Can AI Meet Kids at Their Reading Level?

Reading-Level Probing: what we learned testing whether AI can rewrite real classroom passages to the right grade — across 12 CKLA passages and three AI tools.

Jeremy Maller • Lesley Tompsett • Gabrielle Vandenburg • Summer 2026



Get to Know Us

The teachers behind the research.



Jeremy
Maller

- Coquillard Elementary
- 4th Grade
- 25 Years of Teaching
- Dad to Addie (20), Sam (18), Lauren (16), and Stella (10)



Lesley
Tompsett

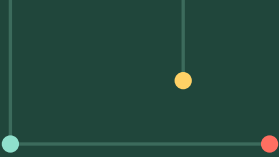
- Nuner Fine Arts Academy
- Pre-K - 5 Visual Art
- 27 Years of Teaching
- Mom to Joey (18)



Gabrielle
Vandenburg

- Nuner Fine Arts Academy
- 5th Grade
- 25 Years of Teaching
- Mom to Grace (22) and Blake (18)

Why We're Digging Into This



1

One class, many levels

Readers in the same room span several grade levels every single day.

2

Leveling by hand takes hours

Rewriting a passage for each reader eats up scarce prep time.

3

AI promises instant rewrites

But can we trust the grade level it claims to hit? That's what we're testing.

The question: when we ask AI to “write this for grade-G,” does it actually land there?

How the Study Works

STEP 1

Start with a real passage

Pull an actual CKLA textbook passage; we already know its grade and its genre.



STEP 2

Ask AI to re-level it

Prompt each tool:
“Differentiate this text to a grade-G level.”



STEP 3

Check the result

Measure the new reading level and confirm the key ideas survived.

Three tools on trial:

ChatGPT

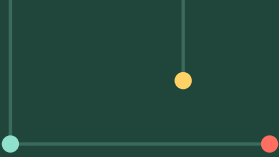
Claude

Gemini

Study Data at a Glance

Passage Details	Passage Text	What we asked AI to do	Which AI is used	Results	Ask Each AI what Grade Level the new result is written in	Rater A	Rater B
4th grade CKLA textbook Fiction This text is written at a 6th to 7th-grade reading level	One day he rescued a child from the path of a moving train, and the boy's father, who was trained to operate a telegraph, offered him lessons in telegraphy. He soon became a telegraph operator and before long was inventing remarkable improvements to the telegraph that got the attention of financiers. With their support, he opened a laboratory in Menlo Park, New Jersey.	Differentiate the following text to a 4th grade level:	ChatGPT	One day, he saved a young boy from being hit by a moving train. The boy's father was a telegraph operator. To thank him, the father taught him how to use a telegraph. He learned quickly and soon became a telegraph operator himself. Before long, he began inventing better ways to use the telegraph. His new ideas impressed people who were willing to help pay for his work. With their support, he opened a laboratory in Menlo Park, New Jersey, where he continued creating new inventions.	ChatGPT: N/A Claude: 4/5 grade reading level Gemini: 3/4 grade reading level	Said the boys father taught the telegraph operations to thank the man for saving his son. (The original passage did not say that.) Said he continued creating new inventions. (Original passage did not say that.)	The added details go beyond the original passage, so I would be cautious about accuracy even though the reading level seems appropriate.
			Claude	One day, he saved a child from a moving train. The boy's father knew how to use a telegraph. He offered to teach the young man telegraphy. The young man soon became a telegraph operator. Before long, he was inventing amazing improvements to the telegraph. These inventions got the attention of rich businessmen called financiers.	ChatGPT: 4 grade reading level Claude: N/A Gemini: 2/3 grade reading level	This passage is more condensed than the other passages. Described financiers as rich businessmen.	This version is shorter and easier to read, and explaining financiers as rich businessmen supports student understanding.
			Gemini	Here is a version of the text differentiated to a 4th-grade reading level. It defines what a telegraph is right inside the story and breaks up the long sentences so a 9- or 10-year-old can easily follow Thomas Edison's journey. One day, Thomas Edison saved a little boy from being hit by a moving train. The boy's father was so grateful that he offered to teach Edison a special skill. The father knew how to use a telegraph. A telegraph was an early machine used to send coded messages over long distances using wires. Edison learned quickly and soon got a job operating telegraph machines. Before long, he started inventing great ways to make the machines work even better. His amazing ideas caught the attention of wealthy businessmen who had a lot of money. With their helpful financial support, Edison was able to open his very own science laboratory in Menlo Park, New Jersey.	ChatGPT: 4/5 grade reading level Claude: 4/5 grade reading level Gemini: N/A	The new passage gives a brief overview of what the change is. The new passage says Thomas Edison was the man who saved the boy. (The original passage did not give his name.) The passage adds that he was able to make the machines better and that he opened his own science lab. The original passage just said he opened a laboratory.	The rewrite explains the change, but it adds details not stated in the original, so it should be checked before classroom use.

How We Score It



Ground rules we settle before any model runs, so every number means the same thing.

● Reading level

Taken straight from the CKLA textbook; we score against the AI's grade level.

● "Hitting" a level

Counts if it lands within one grade either way.

● Runs per task

Three, so we can see how much the answer varies.

● What counts as correct

The right idea: loose wording is fine, wrong answer isn't.

● Two kinds of ask

A plain solve, and "write it for grade G."

● Same source, every time

Real CKLA passages with a known grade and answer.

What We've Tested So Far

12

passages
re-leveled

3

AI tools
compared

2

text types:
fiction + nonfiction

4-5

target grade
band

Cross-checked against the CKLA rubric

Every rewrite is compared back to CKLA's three-part text-complexity rubric: quantitative level, qualitative demands, and how the text is actually taught. **Teacher notes ride along with each passage**, flagging when a rewrite drifts too easy, too hard, or loses the point.

Results at a Glance

The headline numbers across 36 logged trials.

62.5%

Overall accuracy

22 of 36 trials correct

60%

**Level-on-command
success**

21 of 36 “write-for-grade” hits

0.98

Avg. grade-level gap
when a grade was requested

3.8

Max run-to-run spread
grades, single-task spot check

- **Where difficulty bites:** grade-5 multi-step problems were missed in every logged run, while grade-2 and grade-4 tasks were correct every time – suggestive, not conclusive.

Accuracy Drops on the Hardest Grade

Two of three task grades were perfect; the hardest missed every run.

100%

Grade-2 tasks

correct in every run

100%

Grade-4 tasks

correct in every run

0%

Grade-5 tasks

missed in all logged runs

- **The pattern:** the grade-5 task — both fiction and nonfiction — was missed in all logged runs, while grade-2 and grade-4 reading passages were answered correctly every time.

Passage Case Studies

Four moments that stood out across the 12 passages.

PASSAGE 6

Rare agreement

One of the few passages where two models' grade estimates for the same rewrite matched — most passages saw disagreement.

PASSAGE 7

Widest disagreement

Peer grade ratings ranged from 2nd-3rd to 6th-7th grade on one rewrite — the sensitive topic limited what a readability score alone could capture.

PASSAGE 8

Gemini grades low

Teacher noted Gemini consistently rates reading level lower than the other two models when evaluating this passage's rewrites.

PASSAGE 11

Fabrication risk

Both ChatGPT's & Gemini's rewrites added invented details not in the source- naming a historical figure, adding plot points that weren't there.

CKLA Rubric Alignment

Passages 1-12 checked against CKLA's three-part complexity model.

OVERALL JUDGMENT

CKLA generally follows its own rubric well. A few passages score above their nominal grade band, but topic, organization, and expected scaffolding keep them instructionally appropriate.

MAIN RISK IDENTIFIED

Independent readability varies – several 4th/5th-grade passages read as 5th-7th grade by quantitative or AI estimates, so they need teacher support rather than independent reading.

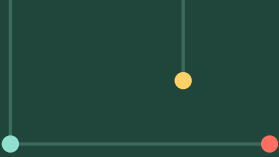
AI REWRITE PATTERN

ChatGPT stays closest to target; Claude often simplifies below it; Gemini tends to preserve or expand detail but sometimes drifts above target.

RECOMMENDATION

Use CKLA originals for teacher-supported instruction, and AI rewrites selectively for access – but first verify vocabulary, sentence length, and concept fidelity (including invented details).

How CKLA Chooses Its Texts



1

Quantitative

Word frequency and sentence length — the Lexile-style measure of raw difficulty.

2

Qualitative

Structure, clarity, knowledge demands, and layers of meaning or purpose.

3

Reader & Task

How the text is taught — read-aloud, scaffolding, vocabulary, and discussion.

The takeaway: a hard text can still be the right text when it is taught with the right support.

The CKLA Gap

1-2

grade levels
above the label

6-7

measured grade,
4th-gr nonfiction

5-6

measured grade,
5th-gr nonfiction

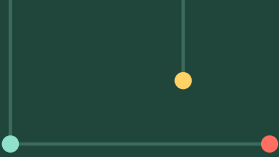
4-5

fiction sits
closest to target

The pattern: nonfiction runs hardest

Several 4th- and 5th-grade textbook passages measured a grade or two above their label — nonfiction most of all. This above-grade difficulty appears **intentional, not accidental**: the texts are built for teacher-supported instruction, not independent reading.

The Three Tools, Head to Head



1

ChatGPT

Most consistent — usually lands near the center of the target grade band.

2

Claude

Simplifies hardest — often drops below the requested grade, sometimes too easy.

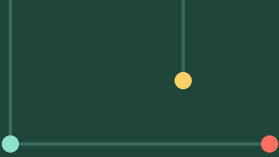
3

Gemini

Keeps the most detail and voice, but often drifts a grade or two above target.

One caution: each tool also graded the others' rewrites — and their grade-level judgments often disagreed.

Can AI Judge Its Own Reading Level?



What we saw

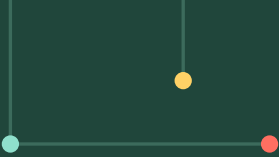
- Grading the same rewrite, the tools often disagreed.
- The same passage was sometimes rated 3+ grades apart.
- Each tool rated its own rewrites more kindly than peers.

Why it matters

- Automated readability scores are estimates, not verdicts.
- No single tool's grade-level claim should stand alone.
- Cross-check with a second measure or a colleague.
- A teacher's read stays the tie-breaker.

AI can draft a reading level — but it can't reliably grade one.

What It Means for Our Classroom



Our recommendation

- Keep CKLA originals for supported instruction.
- Use AI rewrites selectively to widen access.
- Always read a rewrite before it reaches students.

Before you use a rewrite, check

- Reading level and sentence length.
- Key vocabulary and concepts survived.
- Facts stayed accurate, with nothing added.
- Author's voice and cultural context intact.

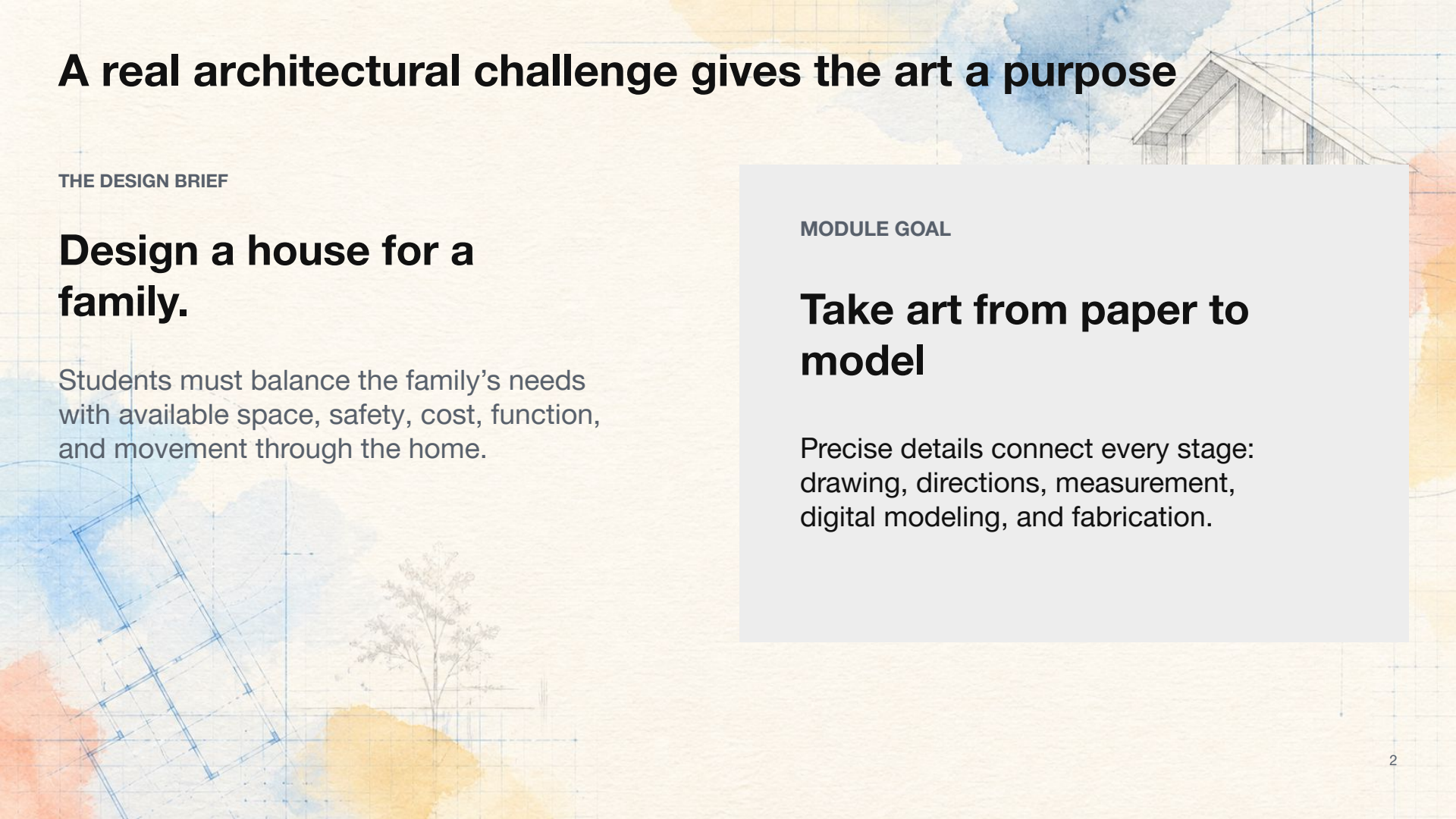
Bottom line: AI can help level a text, but a teacher's eye still decides the grade.

CDI Planning

Ways to Make Art

Students move from a hand-drawn idea to an accurate blueprint, a Tinkercad model, and a 3D-printed house.

A real architectural challenge gives the art a purpose



THE DESIGN BRIEF

Design a house for a family.

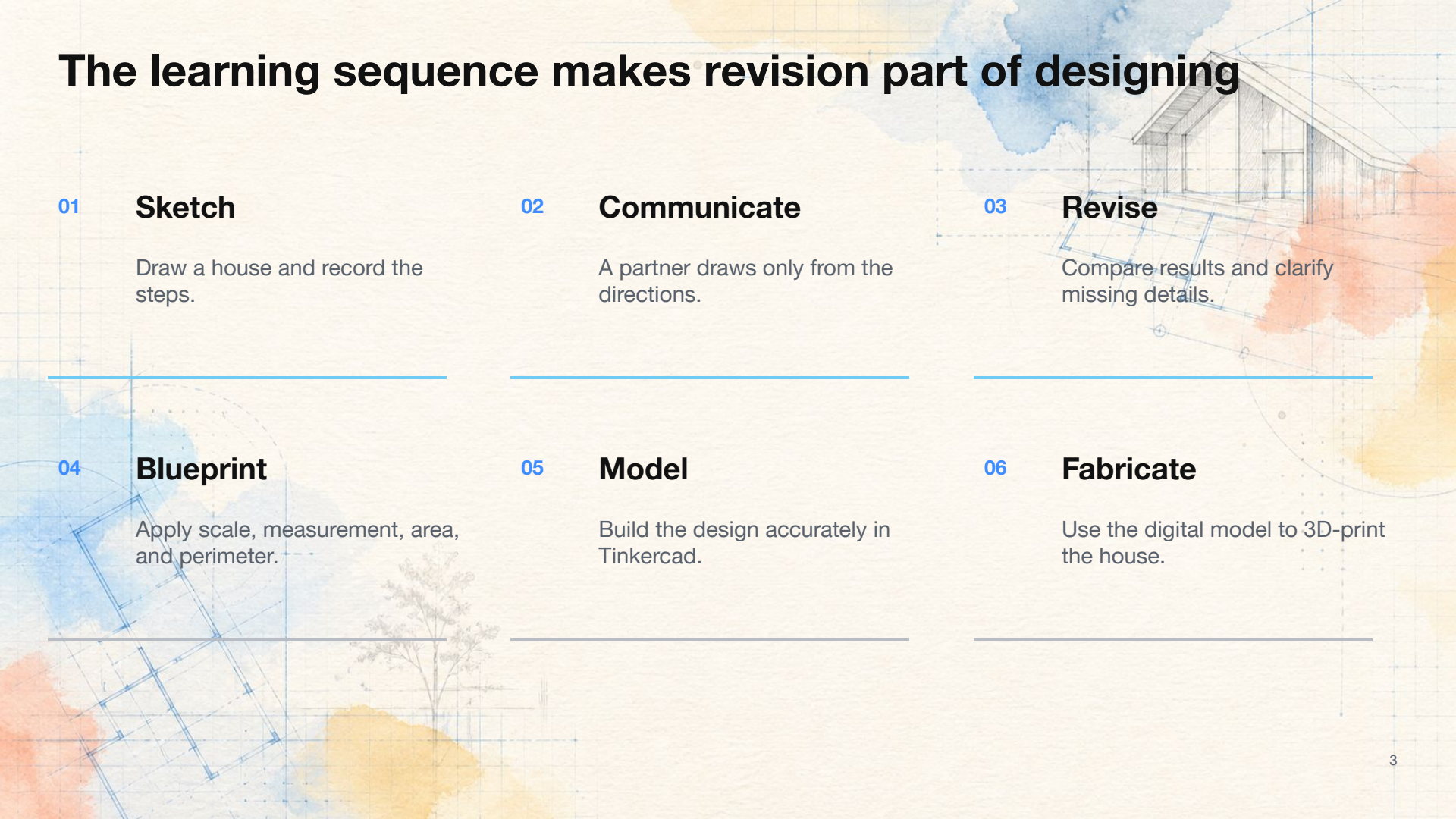
Students must balance the family's needs with available space, safety, cost, function, and movement through the home.

MODULE GOAL

Take art from paper to model

Precise details connect every stage: drawing, directions, measurement, digital modeling, and fabrication.

The learning sequence makes revision part of designing



01

Sketch

Draw a house and record the steps.

02

Communicate

A partner draws only from the directions.

03

Revise

Compare results and clarify missing details.

04

Blueprint

Apply scale, measurement, area, and perimeter.

05

Model

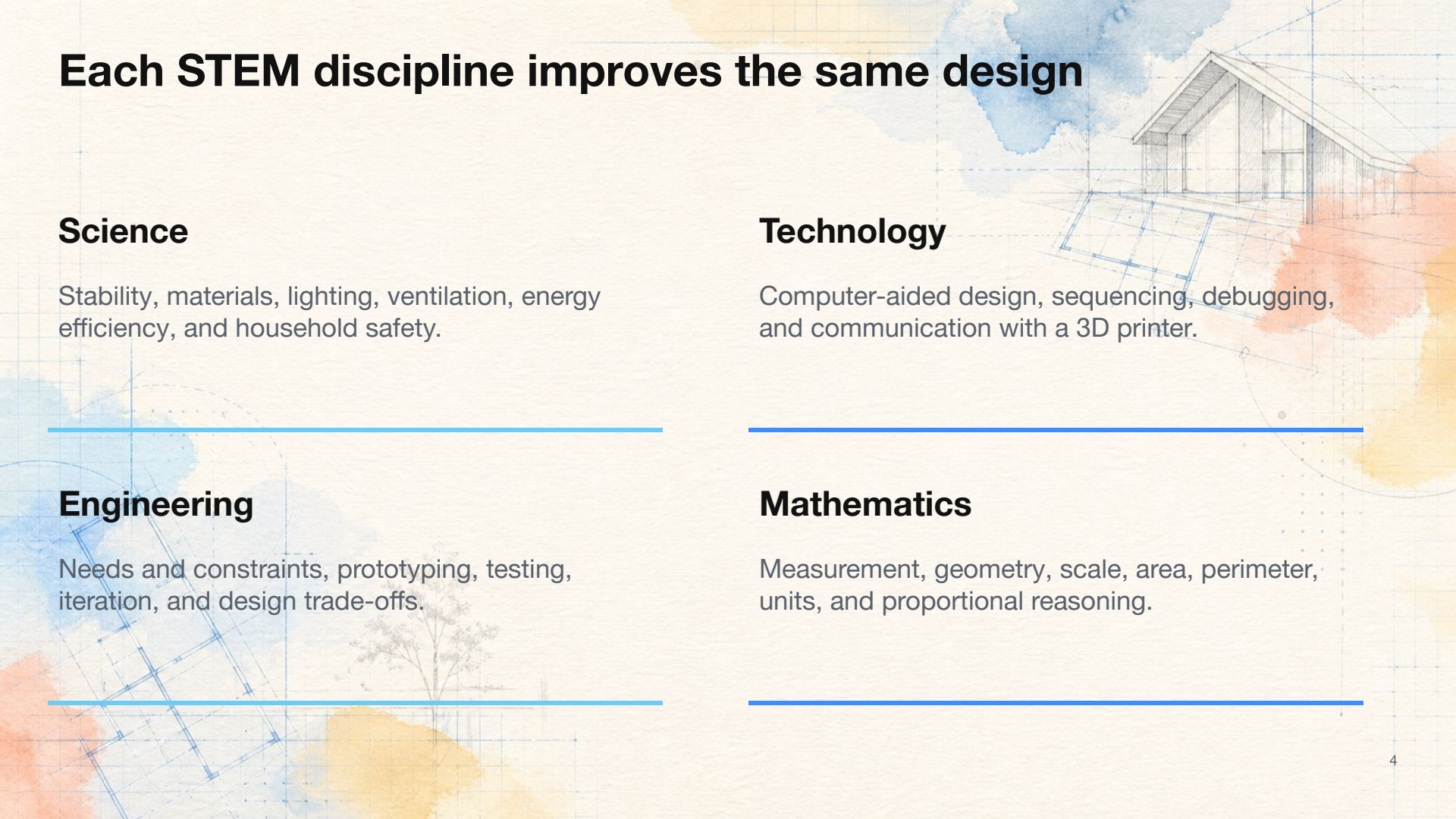
Build the design accurately in Tinkercad.

06

Fabricate

Use the digital model to 3D-print the house.

Each STEM discipline improves the same design



Science

Stability, materials, lighting, ventilation, energy efficiency, and household safety.

Engineering

Needs and constraints, prototyping, testing, iteration, and design trade-offs.

Technology

Computer-aided design, sequencing, debugging, and communication with a 3D printer.

Mathematics

Measurement, geometry, scale, area, perimeter, units, and proportional reasoning.

Four Grade 5 standards anchor the creative work

STANDARD

STUDENT APPLICATION

VA:Cr1.1.5 (a)

Combine ideas to create an innovative house concept.

VA:Cr1.2.5 (a)

Choose among sketching, written procedures, blueprints, and digital modeling.

VA:Cr2.1.5 (a)

Develop skills across multiple art-making techniques through practice.

VA:Cr2.2.5 (a)

Demonstrate craftsmanship and safe use of tools, materials, and equipment.

Assessment happens throughout—not only at the final print

FORMATIVE CHECKS

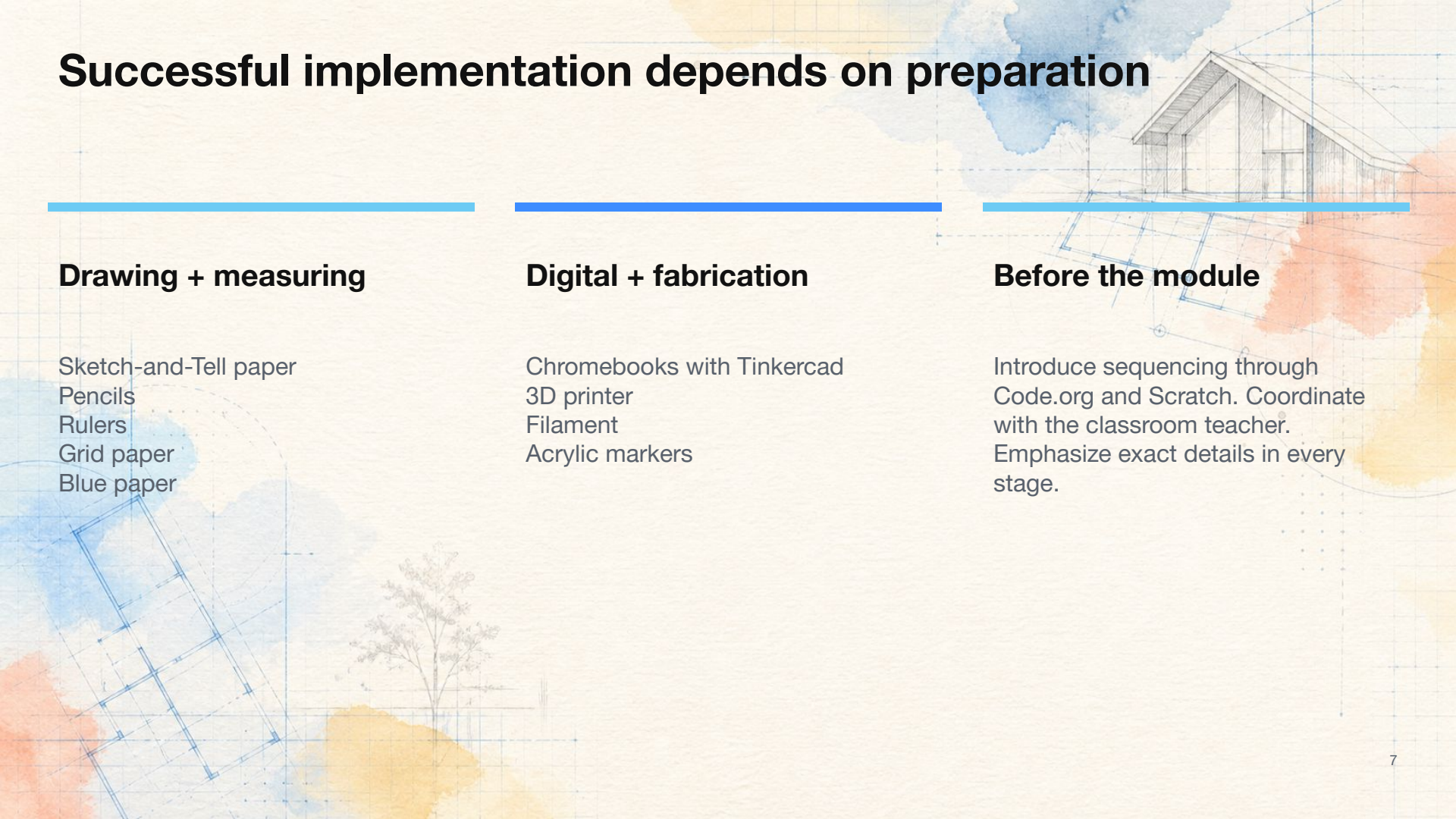
- ✓ Directions are detailed and sequential
- ✓ Partner feedback leads to clearer instructions
- ✓ Blueprint measurements and scale are accurate
- ✓ Tinkercad dimensions and positioning match the plan

SUMMATIVE EVIDENCE

The final house shows whether the full process holds together.

Look for alignment among the sketch, revised directions, blueprint, digital model, craftsmanship, and printed result.

Successful implementation depends on preparation



Drawing + measuring

Sketch-and-Tell paper
Pencils
Rulers
Grid paper
Blue paper

Digital + fabrication

Chromebooks with Tinkercad
3D printer
Filament
Acrylic markers

Before the module

Introduce sequencing through Code.org and Scratch. Coordinate with the classroom teacher. Emphasize exact details in every stage.

Precision turns an idea into something that can be built.

Draw • Explain • Measure • Model • Make

