KidUniversity: ESSA Tier IV ("Demonstrates a Rationale") logic model

Product: KidUniversity (KidU), a K–12 learning app for families, on the web at kiduniversity.net and app.kiduniversity.net
Operator: Arron Pinkney, an individual doing business as KidUniversity · PO Box 155, Edgemoor, SC 29712 · info@kiduniversity.net · (775) 370-9554
Version: October 29, 2026. Every count below comes from the product's own data and code on this date (`node scripts/validate.mjs`, `node website/build.mjs`, `data/standards/coverage.json`).
Status of evidence: KidUniversity has no outcome data yet. No efficacy, correlational or user study has been run. This document makes the research-based case for the design (ESSA Tier IV) and sets out the study that would test it at Tier III. It contains no results, testimonials, user counts, partnerships or awards, because there are none to report.

Contents

1. What Tier IV asks for, and how this document answers it
2. Problem statement and context
3. Target population
4. Inputs and resources
5. Activities (the product's components)
6. Outputs (measurable use)
7. Outcomes: short, medium and long term
8. The logic model on one page
9. The research basis for each link
10. Assumptions and external factors
11. Evaluation plan: a correlational study designed to meet Tier III
12. Limits of this document, and what would strengthen it
13. References

1. What Tier IV asks for, and how this document answers it

The Every Student Succeeds Act defines four levels of evidence (ESEA §8101(21)). The fourth, **"demonstrates a rationale,"** applies to an activity, strategy or intervention that:

1. is based on **high-quality research findings or a positive evaluation** that it is likely to improve student outcomes, set out in a **well-specified logic model**; and
2. includes **ongoing efforts to examine its effects** on those outcomes.

The U.S. Department of Education's non-regulatory guidance, *Using Evidence to Strengthen Education Investments* (first issued 2016, updated 2023), takes its definition of a logic model from EDGAR (34 CFR 77.1): "a well-specified conceptual framework that identifies key components of the proposed process,

product, strategy, or practice ... and describes the relationships among the key components and outcomes." The structure below follows that guidance and the Regional Educational Laboratory workshop toolkit on logic models (Shakman & Rodriguez, 2015): problem, population, inputs, activities, outputs, outcomes, then assumptions, external factors and evaluation.

Tier IV requirement	Where it's met here
A well-specified logic model	§§2–8, with the one-page table in §8
Grounded in high-quality research	§9: each link, with full citations (§13)
Ongoing efforts to examine effects	§11: a planned correlational study designed to meet Tier III ("promising evidence"); §12 says what isn't done yet

Tier IV doesn't require outcome results, and none are claimed.

2. Problem statement and context

The problem. Many students finish a school year without having mastered its core skills, and the gaps add up. On the 2024 National Assessment of Educational Progress, 31% of fourth graders scored at or above NAEP Proficient in reading, and fewer than a third of students in grades 4 and 8 reached that level in reading. A little more than a quarter of eighth graders reached it in math (National Assessment Governing Board, 2025).

What families run into at home. Three problems are common, and KidU is designed around them:

1. **Not knowing where a child stands.** Parents often can't tell which units a child has mastered and which need work, so practice time goes to the wrong things.
2. **Homework help that hands over answers.** When a child is stuck, the fastest help (an older sibling, an answer key or a general-purpose chatbot) often supplies the answer instead of the method. Recent experimental evidence shows that unguarded AI help can raise practice scores while lowering learning (Bastani et al., 2025).
3. **Practice without feedback or review.** Worksheets and many apps mark answers right or wrong but rarely explain the error, bring missed items back later, or check that a skill was kept.

Homeschooling families face all three, and also need a full curriculum, tests and records for their state's rules. **Spanish-speaking families** face the added barrier of materials only in English.

Context. KidU is a direct-to-family product. It isn't sold to schools today, and it isn't a school or an accredited program. Families use it in a web browser; iPhone, iPad and Android apps are coming soon. Unit 1 of every course, the check-up, the study plan and parent reports are free; KidUniversity Plus ($7.99 a month or $59.99 a year on the web, one plan per family) unlocks the rest. There are no ads and no third-party tracking.

3. Target population

Group	Who	How KidU serves them
Primary: learners	Children and teens in kindergarten through grade 12 (about ages 5–18) learning at home: after school, in the summer or as homeschool	EarlyU (K–6, "Learn. Play. Grow.") and TeenU (grades 7–12, "Grow. Build. Become."), a calmer look for teens with a course view and SAT® and ACT® prep
Primary: parents and guardians	The adult who creates the family account (children can't), sets rules and reads reports	Grown-ups area behind a parental gate: reports, weekly email, screen-time rules, consent for AI features, homeschool records
Subgroup: homeschool families	Parents using KidU as all or part of a curriculum	180-day courses, unit tests, a printable homeschool report, an hours and attendance log with CSV export
Subgroup: Spanish-speaking families and English learners	Families who speak Spanish at home	Screens, Professor Hoot's voice and Ask Hoot's answers in Spanish; lessons translated course by course (an untranslated lesson shows in English); more than 1,000 picture books in Spanish
Subgroup: early and struggling readers	K–2 learners and older students who read below grade level	"Read to me" read-aloud with word highlighting; a "Read the question" button in K–1; animated lessons with captions

Not the target today: classroom use run by a teacher. KidU has no teacher accounts or rosters. A school or district evaluation would need those, or a home-use design such as the one in §11.

4. Inputs and resources

Input	What exists today (source in the repo)
Curriculum	**60 full-year courses** across **13 grades (K–12)**: Reading, Writing, Math, Science and Social Studies for K–5; English, math, science and history for grades 6–8; named high school courses (for example Algebra 1, Biology, U.S. History); and SAT and ACT prep. Each course is a **180-day school year**: **9,531 lessons** (9,482 skill lessons and 49 projects), **1,269 unit tests** of at least 15 questions each, in **1,303 units**, with **134,509** scored lesson and test questions (`data/courses/*.json`, `data/skills-seed.json`, `data/lessons/`; counted with `scripts/curriculum.mjs`)
Standards alignment	Grades K–8 math and ELA mapped to the Common Core at the standard level (`docs/ccss-coverage.md`, `data/standards/`). **Math: 227 of 229 standards covered (99%), 2 partly. ELA: 269 of 373 covered (72%), 96 partly, 8 not covered**: reading, language and foundational skills are covered well; writing is practiced with whole pieces in the app's writing activities (A16), and speaking and listening with listening activities and talks with a grown-up. Standards that need publishing, research with outside sources, live discussion or presenting stay partly covered (the rules are in `website/standards-data.mjs`). Grades 2–8 math follows the Illustrative Mathematics sequence (`data/curriculum-map.json`). High school, the literacy standards for history and science, and NGSS aren't mapped yet
Session design	Lesson steps built by `convex/sessionLogic.ts` and `convex/extrasLogic.ts` (§5, A3)
AI tutor	Professor Hoot's step-by-step help, Ask Hoot and My Schoolwork use Claude, by Anthropic, under system prompts in `convex/stepHelpAi.ts`, `convex/hootAi.ts` and `convex/schoolworkAi.ts`

Input	What exists today (source in the repo)
Supplementary content	A read-aloud library of more than 2,700 books (original stories, public-domain classics and openly licensed picture books, more than 1,000 in Spanish); 868 Watch & Learn videos from sources such as NASA and the National Park Service, plus animated Professor Hoot lessons for every unit in K–4; six learning games and the Fluency Zone for math facts
Parent tools	Progress report, weekly email, screen-time rules, homeschool report and hours log (`src/Report.tsx`, `convex/weekly.ts`, `src/ScreenTime.tsx`, `src/Records.tsx`)
Privacy and safety	COPPA measures documented in `docs/coppa-seal-prep.md` and `docs/data-retention-policy.md`: parent-only accounts with email verification, email-plus parental consent before any AI feature, name scrubbing before text reaches the AI, fixed retention periods, in-app deletion
Technology	Convex (database and server functions), Vercel (hosting), Resend (email), Cloudflare Stream (video); web app plus iOS and Android builds
People	One operator (Arron Pinkney), who builds and runs the product. **There's no in-house researcher, learning scientist or certified teacher on staff today** (see §12)
Funding	Subscriptions. No grants or outside investment

5. Activities (the product's components)

Each activity names the design element that the research in §9 supports.

A1. Check-up and study plan (diagnostic placement and formative assessment).
A short placement check for one subject at one grade, about 10 minutes. It asks 3 questions from each of up to 6 units spread across the year (4 units for reading and English, where each unit's questions share one passage), drawn from that grade's own lessons. "I'm not sure yet" is always allowed. A unit is "already strong" at 2 of 3 right. At 75% or more overall, Hoot suggests the next grade's check-up; below 50%, the grade before (`convex/checkupLogic.ts`: `STRONG`, `MOVE_UP`, `MOVE_DOWN`). The plan then orders the work: units with misses first, then the rest of the year, while strong units can be skipped. "Today's plan" shows the next lesson in up to four subjects. A photographed page of schoolwork (A6) can also move a unit into "practice first" or "already strong." "My check-up answers" shows each question with the child's answer and, for a miss, the lesson's explanation and example.

A2. Unit tests and mastery.
Every unit ends with a test of at least 15 questions. **The next unit opens only after 85% or more right on the first try on the unit test** (the mastery gate: `convex/mastery.ts`, `convex/masteryLogic.ts`), enforced on the server for every way into a locked unit's lessons, tests and activities. Below the goal, the learner sees a kind screen with their score, the 85% goal and the lessons the missed questions came closest to ("Practice these, then try again"); those lessons' questions go to "Practice what I missed" for that unit, and the study plan's next step is that review, then the retake. Retakes are unlimited, come in a new order (and draw a different 15 when the test has more questions), and the best score is kept. Parents can set the goal per child to 70, 80, 85 (default) or 90%, and can let a learner move on without it; both need the parent's password, and every "move on" is logged and shown in Grown-ups and the homeschool report. Units a check-up shows are already strong count as passed, so placement never strands a learner, and units a learner had already started before the gate (October 29, 2026) stay open. A grade "graduates" (with a printable diploma) only when every lesson and unit test in every course in that grade is done.

A3. Guided lesson sessions (explicit instruction, worked examples, guided practice, retrieval).
Each lesson is one school day, run as a timed, step-by-step session whose length depends on the grade band (`convex/sessionLogic.ts`, `convex/extrasLogic.ts` `PLAN`):

Band	Target length	Steps, in order
K–1	about 20 minutes	Warm-up (3 review questions) → Learn (explanation and example) → Watch → Practice (about 12 questions) → Play (a short learning game) → Exit check (3)
Grades 2–5	about 30 minutes	Warm-up (4) → Learn (explanation, example and up to 3 worked examples) → Watch → Practice (up to 20) → Exit check (3)
Grades 6–12	about 60 minutes	Warm-up (5) → Read closely (English, reading and history passages) → Learn (explanation, a deeper section and up to 4 worked examples) → Watch → Guided practice (up to 8) → Practice (up to 27) → Challenge (up to 4–5 harder questions) → Exit check (5)

- **Warm-up** questions come from earlier lessons in the unit, or the previous unit for a unit's first lesson, so each session starts with retrieval of earlier material.
- **Exit check** questions are ones not used earlier in the session: the lesson's own exit questions where they've been written, otherwise questions from the unit test or unit review.
- **Feedback:** a wrong answer gets "Not quite. Try again." with the question's hint. After a second wrong try, the learner can open "How to do it" (the lesson's explanation and example) and, with a parent's consent, Professor Hoot's step-by-step help (A4).
- **Scoring:** only the lesson's own questions are scored, by right on the first try. Each earns a star; anything not right on the first try goes to "Practice what I missed" (A5).
- **How lessons are written:** the core lessons, passages and questions are original KidUniversity content. For grades 2–12, the session extras (worked examples, deeper sections, guided, challenge and extra practice questions, exit checks) are written in batches by an AI model (Claude) and checked automatically for format and answer consistency before use (`convex/lessonExtrasBatch.ts`, `checkExtras`). Lessons without extras borrow review questions from the unit.

A4. Professor Hoot's step-by-step help on wrong answers (elaborated feedback, AI tutoring).
After a wrong answer, "Show me step by step" sends the question, the correct answer and the learner's own answer to the AI. It returns (1) one or two kind sentences on what likely went wrong, judged from the answer given; (2) each step from question to answer; (3) the answer, which must match the answer key; and (4) a similar practice question with its answer (`convex/stepHelpAi.ts`). It's available in lessons and in "My check-up answers," not during the check-up itself. It needs Ask Hoot to be turned on with the parent's consent, and there's a daily limit.

A5. "Practice what I missed" (retrieval of missed items, review until correct).
Every lesson question not answered right on the first try is saved. "Practice what I missed" serves them in rounds of 12, oldest first (`convex/practice.ts`). A question leaves the list only when it's answered right on the first try in review. A newer attempt at the same lesson replaces that lesson's list. *Design note:* missed items come back when the learner opens the review, not on a timed, expanding schedule. The warm-ups (A3) provide the scheduled, cumulative review.

A6. My Schoolwork: homework-photo help that teaches instead of answering (scaffolding).
A child photographs or uploads a worksheet, homework, test, notes or a study guide. Professor Hoot reads it, explains the ideas step by step, matches the page to KidU lessons and makes a short practice quiz on the same skills (`convex/schoolworkAi.ts`). The system prompt says: "Teach, don't hand over answers," and

"Never fill in answers to homework or questions that are blank or not yet graded. Instead show the method on a similar example with different numbers or words, and invite the child to try their own." The matched lessons feed the study plan (A1). It's off until a parent consents by email. By default, photos are deleted as soon as Hoot has read them, and always within 30 days.

A7. Ask Hoot (on-demand AI explanations, parent controlled).
A learner types a question, optionally choosing a subject, and gets a kid-friendly, step-by-step explanation. It's off until a parent turns it on and confirms by email; parents can read, report and delete every question; answers are described to families as written by an AI that can make mistakes; and questions are kept 90 days. Free accounts get 10 questions a day.

A8. Read-aloud and the Library (supported reading, vocabulary).
"Read to me" reads lessons and questions aloud. In the Library, more than 2,700 books (including more than 1,000 Spanish picture books) are read aloud with each word highlighted as it's read. K–1 questions have a large "Read the question" button.

A9. Watch & Learn (multimedia explanation).
Animated lessons with Professor Hoot for every unit in K–4, with captions, and 868 curated videos for grades 5–12. A "Watch" step sits in the lesson session when a unit has video.

A10. Fluency Zone and learning games (automaticity through timed practice).
60-second rounds on sets of addition, subtraction, multiplication (0–12 tables) and division facts, tracking each fact as new, tricky, learning or mastered (`convex/fluencyLogic.ts`). Six games (Number Line Drop, Balloon Pop, Sort It!, Match Up, Sentence Builder, Spell It) practice math facts, spelling, words and sentences.

A11. Parent reports, weekly email and screen-time rules (parent involvement and monitoring).

- **Progress report:** lessons this week, days learning, streak, minutes learned, share right on the first try by subject, books finished, and "worth another look" (lessons under 70% right on the first try).
- **Weekly email** every Sunday, in the app's language, which parents can turn off.
- **Screen time:** a daily limit from 15 minutes to 2 hours, "Finish today's plan first" (games and videos wait until today's lessons are done) and quiet hours.

A12. Homeschool records.
A printable report for each child and school year (courses, units and lessons finished, average unit-test scores, check-ups, days and hours, with a signature line) and an hours and attendance log that counts time in the app by day, takes the parent's own entries and downloads as CSV.

A13. English and Spanish.
Each child's app can switch between English and Spanish at any time: screens, Professor Hoot's voice and Ask Hoot's answers. Lessons are translated course by course; an untranslated lesson shows in English.

A14. Motivation and routine.
A star for every question right on the first try, daily streaks, stickers and badges, and a graduation ceremony with a personalized diploma at the end of each grade.

A15. Privacy, safety and accessibility (enabling conditions).
No ads, no ad networks, no third-party analytics or tracking, nothing for children to buy, and no messaging between users. Children can't create accounts. AI features stay off until a parent consents by email. Learner names are replaced with "the student" before any text goes to the AI. Retention periods are fixed (`docs/data-retention-policy.md`). Accessibility features include strong contrast, large tap targets, screen-

reader labels, read-aloud, captions, sound off and reduced motion on the website. No outside accessibility audit has been done yet.

A16. Writing, listening and speaking (writing process, feedback, oral language).
Each English unit (Writing in K–5) has a "Writing & speaking" step with up to three activities (`data/writing/*.json`, `convex/writing.ts`): a guided writing piece (a prompt with read-aloud; plan, draft, revise and edit steps; a word goal; a word bank; a checklist the child ticks; autosaved), a listening activity (a passage read aloud, hidden text that can be shown, then questions), and a speaking activity (a talk with a grown-up, who marks it done; nothing is recorded). With the parent's Ask Hoot consent, Professor Hoot gives 2–3 specific suggestions tied to the checklist and one question to help revise, and never rewrites the piece (`convex/writingAi.ts`). The activities are encouraged, not required for the mastery gate. Parents read every piece in Grown-ups → Writing and can add finished writing to the homeschool report.

6. Outputs (measurable use)

The app already records each of these per learner; §11 explains how they'd be collected for a study.

Output	Measure	Where it's recorded
Placement	Check-ups taken per subject and grade; score; units strong vs. to practice	`checkups`
Lessons	Lessons finished; share right on the first try	`progress`
Dosage	Minutes in lesson sessions per day and week; learning days	`lessonSessions` (seconds and day log), `days`, `usage`
Mastery	Unit tests attempted and passed (≥ the learner's goal, 85% by default); tries and best score; parent overrides	`testScores`, `progress` (`test: true`, `goal`), `masteryOverrides`
Review	Missed questions added, and cleared on a first-try fix	`misses`
Help	Step-by-step requests; Ask Hoot questions; schoolwork uploads (only where a parent consented)	`stepHelp`, `hootChats`, `uploads`
Reading and video	Books started and finished; videos watched	library and video tables
Fluency	60-second rounds; facts mastered	fluency tables
Parent engagement	Reports opened; weekly email on or off; screen-time rules set; homeschool report printed	settings tables (report opens would need a new count; see §11)
Language	Learners using Spanish	per-learner language setting
Writing, listening and speaking	Pieces started and finished; words; checklist; listening scores; speaking marked done; Hoot tips requested	`writingPieces`

7. Outcomes

These are the outcomes the design is **intended** to produce. None has been measured yet.

Short term (within a unit, about 2–6 weeks of regular use)

- S1. Learners and parents know which units are strong and which need work (check-up and plan).

- S2. Learners correct specific errors: missed questions are fixed on a later first try (A4, A5).
- S3. Learners answer more lesson questions right on the first try as they work through a unit's lessons.
- S4. Learners attempt homework with a method instead of copying an answer (A6).
- S5. Parents check progress and set routines (reports, weekly email, screen-time rules).

Medium term (a semester or school year)

- M1. Unit mastery: more unit tests passed at ≥85%, with fewer retakes needed over time and few parent overrides.
- M2. Growth on repeated check-ups: higher scores at the same grade, or placement at the next grade.
- M3. Retention: exit-check and warm-up accuracy on earlier material holds up.
- M4. Math-fact fluency: more facts mastered and more correct answers per 60-second round.
- M5. Steady learning routines: more learning days a week and more of the planned lessons done.
- M6. For Spanish-speaking families: as much use and progress as English-speaking families.

Long term (one or more years)

- L1. Grade-level mastery in core subjects, shown by completing a grade's courses and by external measures (state tests or nationally normed tests the family already takes).
- L2. Closed learning gaps: students who started below grade level on the check-up reach grade level.
- L3. Independent learners who use help to learn a method rather than get an answer.
- L4. For homeschool families: a complete, documented school year that meets their own records needs.

---

8. The logic model on one page

Inputs	Activities	Outputs	Short-term outcomes	Medium-term outcomes	Long-term outcomes
60 full-year courses (K–12 plus SAT/ACT prep): 9,531 lessons, 1,269 unit tests, 134,509 questions	**A1** Check-up and study plan	Check-ups taken; units placed "strong" or "practice first"	**S1** Learner and parent know what to work on	**M2** Higher check-up scores, or placement at the next grade	**L1** Grade-level mastery in core subjects
Common Core mapping: math 99%, ELA 72% (K–8)	**A3** Guided sessions: warm-up, learn, worked examples, guided practice, challenge, exit check	Lessons done; minutes; learning days; first-try accuracy	**S3** First-try accuracy rises across a unit	**M3** Earlier material retained (warm-up and exit-check accuracy)	**L2** Gaps closed for learners who started below grade level
Session design (`sessionLogic.ts`)	**A2** Unit tests (85% opens the next unit; review, unlimited retakes, parent override)	Tests attempted and passed; retakes; overrides	—	**M1** More unit tests passed at ≥85%	
AI tutor (Claude) with teach-don't-tell prompts; parent consent	**A4** Step-by-step help on wrong answers · **A6** My Schoolwork · **A7** Ask Hoot	Help requests; uploads; similar-question attempts	**S2** Errors corrected · **S4** Method, not copied answers		**L3** Learners who use help to learn
Missed-question store	**A5** "Practice what I missed"	Misses added and cleared	**S2**	**M3**	
2,700+ books; 868 videos and K–4 animations; games	**A8** Read-aloud and Library · **A9** Watch & Learn · **A10** Fluency Zone and games	Books finished; videos; fluency rounds; facts mastered		**M4** Math-fact fluency	
Parent tools; Spanish; privacy and COPPA safeguards; one operator	**A11** Reports, weekly email, screen time · **A12** Homeschool records · **A13** English/Spanish · **A14** Stars and streaks · **A15** Privacy and accessibility	Reports viewed; rules set; homeschool reports; Spanish use	**S5** Parents monitor and set routines	**M5** Regular routines · **M6** Equal use and progress in Spanish	**L4** Documented homeschool year

Assumptions: families use KidU regularly (about one lesson per subject on school days); lesson content is accurate and on grade level; parents read reports. *External factors:* school instruction, home language and time, device and internet access, and the child's needs. See §10.

9. The research basis for each link

Each row connects a component to the outcome it's meant to produce, and names the research that makes the link plausible. These studies were done in other settings; they support the design, not a claim that KidU

itself works.

Link	Research that supports it	How KidU applies it, and its limits
A1 → S1, M2. Assess first, then target instruction to what isn't mastered yet	**Formative assessment:** Black & Wiliam (1998) reviewed classroom studies and found that strengthening the feedback students get about their learning yields substantial gains. **Mastery learning:** Bloom (1984); Guskey (2010); Kulik, Kulik & Bangert-Drowns (1990), a meta-analysis of 108 controlled evaluations that found positive effects on exam performance	The check-up places each unit as "strong" or "practice first," and the plan sends practice to gaps first. It's a short screener (3 questions a unit), not a validated diagnostic; reliability is one of the study's research questions (§11, RQ4)
A2 → M1. A mastery threshold with corrective work and retests	Bloom (1984); Guskey (2010); Kulik et al. (1990)	An 85% first-try mastery goal that opens the next unit, corrective work aimed at the lessons behind the missed questions, unlimited retests in a new order with the best score kept, and graduation only when all lessons and tests are done. As in classic mastery learning, the learner doesn't move on before reaching the goal, unless a parent lowers the goal or overrides it (logged)
A3 (warm-up, exit check), A5 → S2, M3. Retrieval practice: answering questions strengthens memory more than rereading	Roediger & Karpicke (2006): practice tests produced better delayed retention than restudying. Dunlosky et al. (2013) rated practice testing as having **high utility**	Every session opens and closes with questions on earlier material. Missed items return in "Practice what I missed"
A3 (warm-ups across the unit), A5 → M3. Spaced (distributed) practice	Cepeda et al. (2006), a quantitative synthesis of 317 experiments, found spacing improves retention and that the best gap grows with the retention interval. Dunlosky et al. (2013) rated distributed practice **high utility**	Warm-ups revisit earlier lessons over days; exit checks draw on the unit test. **Limit:** missed questions aren't on an expanding schedule (§12)
A3 (mixed review) → M3. Interleaving problem types	Rohrer, Dedrick & Stershic (2015): interleaved math practice beat blocked practice on immediate and delayed tests in a classroom study	Warm-ups and extra practice mix questions from several lessons in the unit. The practice step itself is mostly blocked on the day's skill
A3 (learn, worked examples, guided practice) → S3. Worked examples and explicit instruction for novices	Sweller (1988) on cognitive load during problem solving; Atkinson, Derry, Renkl & Wortham (2000) on design principles from worked-example research	An explanation and example, then up to 3–4 worked examples (grades 2–12), then guided practice (grades 6–12) before independent practice: a move from studying examples to solving problems
A3 (hints), A4 → S2, S3. Immediate, elaborated feedback that explains the error	Shute (2008), a review of formative feedback; Hattie & Timperley (2007), a model of feedback built on meta-analyses, which finds the type of feedback matters	A hint on the first miss, then "How to do it," then step-by-step help that names the likely mistake from the learner's own answer, shows each step and gives a similar question. The help is withheld during the check-up so it stays a clean measure

Link	Research that supports it	How KidU applies it, and its limits
A4, A7 → S2, L3. Step-based tutoring by computer	VanLehn (2011): step-based intelligent tutoring systems were nearly as effective as human tutoring (effect sizes about 0.76 vs. 0.79). Kulik & Fletcher (2016), a meta-analysis of 50 controlled evaluations, found a median gain of 0.66 standard deviations	Hoot's help is step-based. **Limit:** these studies tested purpose-built tutors with expert-built models, not general large language models. Whether an LLM tutor matches them is open; §11 measures it
A6 → S4, L3. Scaffold homework rather than supplying answers	Wood, Bruner & Ross (1976) introduced "scaffolding," help that lets the learner do as much as possible alone. Koedinger & Aleven (2007) describe the "assistance dilemma" of when to give and when to withhold help. Bastani et al. (2025): in a field experiment with about 1,000 high school students, an unguarded GPT-4 tutor raised practice scores but lowered later exam scores; a version with learning safeguards largely avoided the harm. Cooper, Robinson & Patall (2006): homework generally relates positively to achievement	Hoot never fills in blank or ungraded homework answers; it shows the method on a similar example, then gives a practice quiz on the same skills. This is the safeguard Bastani et al. found mattered
A8 → S3 (reading), L1. Read-aloud and interactive book reading	Mol, Bus & de Jong (2009): interactive book reading had a moderate effect on oral language in a meta-analysis of 31 studies. Wood, Moxley, Tighe & Wagner (2018): text-to-speech and read-aloud tools improved reading comprehension for students with reading disabilities (average effect 0.35, 22 studies)	Word-highlighted read-aloud in the Library and "Read to me" in lessons. KidU's reading is with an app rather than an adult, so the interactive-reading research is only partly applicable; the parents guide asks families to read together
A9 → S3. Learning from words and pictures together	Mayer (2009), *Multimedia Learning*, sets out research-based principles for combining narration and graphics	Short animated lessons with narration and captions in K–4; curated video in grades 5–12, placed inside the session before practice
A10 → M4. Timed practice of basic facts	Codding, Burns & Lukito (2011): a meta-analysis of basic-fact fluency interventions found drill and practice with modeling produced the largest effects	60-second rounds, five-star steps and per-fact mastery tracking
A11, A12 → S5, M5. Parent involvement and monitoring	Wilder (2014), a synthesis of nine meta-analyses, found a positive relationship between parental involvement and achievement across definitions and measures. Jeynes (2012): parental involvement programs had an overall effect of about 0.3 standard deviations in 51 studies of urban students	Reports, a weekly email, "worth another look" and screen-time rules give parents specific things to see and do. KidU supplies information; it can't make parents use it
A13 → M6. Native-language support for Spanish-speaking learners	Slavin & Cheung (2005), a best-evidence synthesis on the language of reading instruction for English learners, found that most studies of Spanish-dominant elementary students favored bilingual approaches (weighted effect +0.33)	A switch for each child; Spanish explanations keep the English words being learned in English. KidU isn't a bilingual education program, and lessons are still being translated

10. Assumptions and external factors

Assumptions (conditions the model depends on, which KidU partly controls)

1. **Dosage.** Gains depend on regular use, about one lesson per subject on school days. A course is 180 lessons.

2. **Accurate content.** Lessons, answer keys and AI-written extras are correct and on grade level. Extras are checked automatically for format and answer consistency, but there's no documented human review of every item.
3. **AI help is accurate and stays in role.** Hoot's step-by-step answer must match the answer key, but the explanation of the steps can still contain mistakes. Kids can report any answer, and the operator reviews reports.
4. **The check-up places learners well enough.** Three questions a unit is a short sample.
5. **Parents engage.** Reports and rules only help if a parent reads and uses them.
6. **Learners use help to learn.** The design withholds homework answers, but a determined learner can still look for answers elsewhere.

External factors (outside KidU's control)

- School instruction, tutoring and other programs the child uses at the same time
- Family time, home language, parents' education and income
- Device and internet access, and whether a parent can afford Plus beyond the free Unit 1
- The child's prior achievement, disabilities and motivation
- State homeschool rules, which differ widely
- Seasonal patterns (summer, holidays, test seasons)
- Changes in the AI model provider's behavior or availability

---

11. Evaluation plan: a correlational study designed to meet Tier III

ESSA's Tier III ("promising evidence") requires at least one **well-designed and well-implemented correlational study with statistical controls for selection bias** that finds a statistically significant favorable effect on a relevant outcome, with no strong contrary evidence. The study below is designed to that standard. It is a **plan**: none of it has started.

11.1 Who runs it

An independent researcher should design, run and report the study, for example a university faculty member or an evaluation firm, under an agreement that lets them publish whatever they find. The operator supplies de-identified data and doesn't analyze outcomes. An institutional review board (IRB) must approve it before any data are collected.

11.2 Research questions

- **RQ1 (dosage and mastery).** Among learners who take a check-up, is more KidU use (lessons finished, minutes in sessions) associated with more unit tests passed and higher end-of-period check-up scores, after controlling for the starting check-up score and other covariates?
- **RQ2 (feedback and review).** Among learners with similar use and starting scores, is using "Practice what I missed" and step-by-step help associated with higher first-try accuracy on later questions from the same lessons and units?
- **RQ3 (subgroups).** Do the associations in RQ1 differ by grade band (K–1, 2–5, 6–12), subject, app language (English or Spanish) and starting level (below, at or above grade on the first check-up)?

- **RQ4 (measurement).** How consistent are check-up results (agreement between two check-ups taken close together), and how well do they predict unit-test results? This checks whether the check-up can serve as a study measure.
- **RQ5 (external outcome, if available).** For families who share them with consent, are KidU use and mastery associated with scores on an external test the child already takes (a state test, or a nationally normed test such as those homeschool families often use)?

11.3 Design

A **longitudinal correlational study over one school year (about 9 months)**, comparing learners with different amounts of use, with statistical controls for selection:

- **Primary outcome:** change in check-up score at the starting grade in **math and reading/English** from the first check-up (baseline) to a check-up at the end of the period. (A re-check prompt at about 9 weeks and at the end would need to be added to the app; it doesn't exist today.) Taking the end-of-period check-up at the **same grade as baseline** keeps the scale comparable.
- **Secondary outcomes:** unit tests passed; first-try accuracy on unit tests; retained accuracy on warm-up and exit-check items; fluency facts mastered; RQ5's external scores where available.
- **Exposure:** lessons finished and minutes in lesson sessions (dosage), and use of "Practice what I missed," step-by-step help and My Schoolwork.
- **Controls for selection bias:** baseline check-up score (the key covariate), grade, subject, language, homeschool status (parent-reported), Plus status, time of year started, and prior-period use. Methods: **multilevel regression** (learners within families), plus **propensity-score weighting or matching** comparing high-use and low-use learners with similar baseline scores and characteristics. **Sensitivity analyses:** E-values or similar bounds for unmeasured confounding, and results with and without learners whose use stops early.
- **Baseline equivalence:** for any high-use vs. low-use comparison, report baseline differences on the check-up and covariates, aiming for small differences after adjustment (the researcher should confirm the threshold Tier III reviewers apply; 0.25 standard deviations is a common limit).
- **A stronger option, if feasible later:** a randomized or staggered rollout of one component (for example, step-by-step help turned on for a randomly chosen half of consenting families for a term) would move part of the evidence toward Tier II or Tier I.

11.4 Sample

- **Who:** learners in grades 2–8 (where the curriculum and standards mapping are most complete) whose parent opts in to the study, who take a baseline check-up in math or reading/English and use KidU for at least 4 weeks.
- **Size:** set by the researcher's power analysis. As a planning figure, detecting a small association (about 0.2 SD) with typical covariate adjustment and attrition usually needs several hundred learners per grade band. **KidU doesn't have a user base of that size today**, so the study starts when enough families have opted in, or begins with a smaller pilot.
- **Attrition:** report overall and differential attrition; analyze all who meet the minimum, with sensitivity checks.

11.5 Measures and data

Measure	Source	Notes
Check-up score (baseline and end)	`checkups`	Reliability checked in RQ4. Items come from each grade's own lessons
Unit-test pass and score	`testScores` (every try, best score), `progress` (`test: true`)	≥15 questions a test; 85% pass by default (70/80/90 if a parent chose); overrides in `masteryOverrides`
First-try accuracy, lessons and review	`progress`, `misses`	
Dosage	`lessonSessions`, `days`, `usage`	Minutes per day and week
Help use	`stepHelp`, `hootChats`, `uploads`	**Counts only.** No child-written text leaves the system for the study
Fluency	Fluency tables	Facts mastered, correct per round
Covariates	Learner grade, language, Plus status; a short parent survey (homeschool, other programs, external test scores if they want to share)	Survey by opt-in only
Implementation fidelity	Share of planned session steps completed; share of "today's plan" done	From session logs

11.6 Privacy, consent and COPPA

KidU's current privacy policy doesn't cover research use. Before any study data are collected:

1. **Update the privacy policy** (English and Spanish) to describe the study and raise `POLICY_VERSION`, so Grown-ups shows the change.
2. **Separate, opt-in parental consent** for the study, through the same email consent flow used for the AI features (`convex/consent.ts`), with **age-appropriate assent** from children old enough to give it, as the IRB requires. Not taking part changes nothing in the app.
3. **De-identified data only.** The researcher gets a dataset with random study IDs in place of account and learner IDs; no names, emails, typed questions, photos or free text; dates shifted or reduced to week numbers; small groups suppressed in reports (for example, no cell under 10).
4. **Aggregate reporting only.** No individual child is identifiable in any report.
5. **A data-use agreement** with the researcher: purpose limits, security, no re-identification, deletion at the end of the study, and no sale or other use.
6. **Right to withdraw.** A parent can withdraw at any time in Grown-ups, and that learner's data leaves the study dataset.
7. **Retention.** Study data are kept by the researcher only for the agreed period. KidU's own retention schedule (`docs/data-retention-policy.md`) is unchanged.

Have a privacy lawyer review the consent text and the data-use agreement, as `docs/coppa-seal-prep.md` already recommends for the AI consent.

11.7 Analysis and reporting

- A pre-registered analysis plan (for example, on the Open Science Framework or the Registry of Efficacy and Effectiveness Studies) before outcome data are looked at.

- Report effect sizes with confidence intervals, baseline equivalence, attrition and every pre-registered outcome, including null and unfavorable results.
- Write the report to the standards Tier III reviewers apply (§11.3) so it can be submitted to clearinghouses and certifiers.

11.8 Timeline (from the date a researcher is engaged)

Months	Step
0–2	Agree scope with the researcher; IRB application; pre-register; update privacy policy and consent flow; build the de-identified export and the end-of-period re-check prompt
2–3	Opt-in recruitment begins; baseline check-ups
3–12	Data collection over one school year; mid-year fidelity check
12–14	Analysis and report
14–15	Publish the report; submit it for Tier III review (see `docs/accreditation-submission-guide.md`)

11.9 Ongoing monitoring before the study

Until the study runs, the operator can track aggregate, non-identifying usage counts already produced by `npm run stats` (lessons, check-ups, days) to see whether use is high enough for the study's dosage to vary meaningfully. These counts aren't outcome evidence and shouldn't be presented as such.

---

12. Limits of this document, and what would strengthen it

- **No outcome data.** Everything in §7 is intended, not observed.
- **Written by the operator, not by a researcher.** Certifiers and districts give more weight to a logic model reviewed or co-authored by someone with research training. **Recommended:** have an independent learning scientist or education researcher review this document, check the citations and co-sign it before it's submitted anywhere.
- **No qualitative user research yet** (interviews, observations or usability testing with families). Digital Promise's certification requires one such study (see `docs/digital-promise-research-based-design.md`).
- **Spacing is partial.** Missed questions return when the learner opens them, not on an expanding schedule. Scheduling reviews at growing intervals (for example 1, 3, 7 and 21 days) would match the spacing research more closely.
- **Interleaving is partial.** The core practice step is blocked on one skill; only warm-ups and extra practice mix lessons.
- **The mastery gate can be overridden.** The next unit opens at 85%, but a parent can lower the goal or let a learner move on (logged and reported), and learners keep units they'd started before the gate. The missed-lesson review is matched by wording, not by a question-to-lesson key, so it can point at a neighboring lesson.
- **AI-written content.** Extras for grades 2–12 are AI-written and machine-checked; a documented human review of a sample would strengthen the accuracy assumption.
- **Standards mapping stops at grade 8**, and ELA coverage is 72% full (98% full or partial). High school isn't mapped standard by standard (its writing activities list grade-band W and SL standards).

- **No teacher or classroom model**, so the study is designed for home use.

13. References

All citations below were checked against publisher, PubMed, ERIC or university records on October 29, 2026.

- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181–214.
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16. https://doi.org/10.3102/0013189X013006004
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354
- Codding, R. S., Burns, M. K., & Lukito, G. (2011). Meta-analysis of mathematic basic-fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice, 26*(1), 36–47.
- Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research, 76*(1), 1–62. https://doi.org/10.3102/00346543076001001
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266
- Guskey, T. R. (2010). Lessons of mastery learning. *Educational Leadership, 68*(2), 52–57.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487
- Jeynes, W. H. (2012). A meta-analysis of the efficacy of different types of parental involvement programs for urban students. *Urban Education, 47*(4), 706–742. https://doi.org/10.1177/0042085912445643
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239–264. https://doi.org/10.1007/s10648-007-9049-0
- Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60*(2), 265–299. https://doi.org/10.3102/00346543060002265
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78. https://doi.org/10.3102/0034654315581420
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research, 79*(2), 979–1007. ht

tps://doi.org/10.3102/0034654309332561

- National Assessment Governing Board. (2025). *Nation's Report Card: 2024 NAEP reading and mathematics results, grades 4 and 8* [Press release and results]. https://www.nationsreportcard.gov/
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x
- Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology, 107*(3), 900–908. https://doi.org/10.1037/edu0000001
- Shakman, K., & Rodriguez, S. M. (2015). *Logic models for program design, implementation, and evaluation: Workshop toolkit* (REL 2015–057). U.S. Department of Education, Institute of Education Sciences, Regional Educational Laboratory Northeast & Islands.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795
- Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research, 75*(2), 247–284. https://doi.org/10.3102/00346543075002247
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- U.S. Department of Education. (2023). *Non-regulatory guidance: Using evidence to strengthen education investments*. https://www.ed.gov/media/document/non-regulatory-guidance-using-evidence-strengthen-education-investments-2023-46305.pdf
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369
- Wilder, S. (2014). Effects of parental involvement on academic achievement: A meta-synthesis. *Educational Review, 66*(3), 377–397. https://doi.org/10.1080/00131911.2013.780009
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x
- Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis. *Journal of Learning Disabilities, 51*(1), 73–84. https://doi.org/10.1177/0022219416688170