Adaptive Math Pathways (AMP)

Logic model and theory of change

Skylar Saveland

Corpora Inc. · skylar@corpora.inc

Version 1.2 · November 2026 · Software: AMP at commit 7fd3147f0 of github.com/corpora-inc/encorpora

Summary. AMP is an adaptive mathematics program for kindergarten through precalculus. This document sets out the rationale the U.S. Department of Education asks for under the fourth ESSA evidence level: a logic model, the research that informs each of its links, the outcomes it predicts and how each will be measured, and the evaluation that is planned. Every statement about what AMP does cites its public source code at one commit. Every statement about learning science cites published research. The rationale predicts that AMP raises mathematics achievement on an independent, nationally normed test, and that the gain grows with regular use. A phased program of pre-registered studies, from a pilot study to randomized trials, tests that prediction (Section 7).

1 Purpose and the Tier 4 standard

The Every Student Succeeds Act (ESSA) defines an "evidence-based" activity at four levels. The fourth applies to an activity whose effects have not yet been estimated but which rests on research and is being studied. The statute, ESEA section 8101(21)(A)(ii), reads (U.S. Code, 2015):

> "(ii)(I) demonstrates a rationale based on high-quality research findings or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes or other relevant outcomes; and (II) includes ongoing efforts to examine the effects of such activity, strategy, or intervention."

The Department's current regulations (34 CFR 77.1, as amended in August 2024) define the level and the logic model it refers to (U.S. Department of Education, 2024):

> "*Demonstrates a rationale* means that there is a key project component included in the project's logic model that is supported by citations of high-quality research or evaluation findings that suggest that the project component is likely to significantly improve relevant outcomes."
>
> "*Logic model* (also referred to as a theory of action) means a framework that identifies key project components of the proposed project (i.e., the active 'ingredients' that are hypothesized to be critical to achieving the relevant outcomes) and describes the theoretical and operational relationships among the key project components and relevant outcomes."

The Department's non-regulatory guidance, revised in September 2023, summarizes the level as a "logic model informed by research or evaluation findings" (Table 1) and asks that implementation plans include "a logic model, informed by high-quality evidence, that suggests how the project component is likely to improve relevant outcomes" (U.S. Department of Education, 2023, pp. 9, 16).

The archived 2016 edition, which many state documents still reproduce, states the two requirements most concretely (U.S. Department of Education, 2016, p. 9):

> "To demonstrate a rationale, the intervention should include: 1) A well-specified logic model that is informed by research or an evaluation that suggests how the intervention is likely to improve relevant outcomes; and 2) An effort to study the effects of the intervention, ideally producing promising evidence or higher, that will happen as part of the intervention or is underway elsewhere …"

A Regional Educational Laboratory handout puts it plainly: "an effort to study the effects of the program must already be planned or under way" (Regional Educational Laboratory Midwest, 2019). This document is written to the strictest of these texts. It names AMP's key components (Section 4), cites published research for each, states the outcomes each is expected to improve and how they will be measured (Section 5), and describes the evaluation that will examine its effects (Section 7). The logic model uses the components taught in the Regional Educational Laboratory toolkits: resources or inputs, activities, outputs, and short-, mid- and long-term outcomes, with the assumptions behind them (Kekahio et al., 2014; Lawton et al., 2014; Shakman and Rodriguez, 2015). Table 5 maps each criterion to the section that meets it, and Section A sets out every published criterion verbatim beside the place where it is met.

This report is one of the AMP Research series. The phased research program is set out in [RR-1]. AMP itself is described in the Technical Notes: the system [TN-1]; the learner model, placement and national norms [TN-2]; the curriculum, lessons, films and workbooks [TN-3]; the measurement of use [TN-4]; and open research questions [TN-5].

2 The problem and the program

The problem. Mathematics achievement in the United States remains below its pre-pandemic level. In the 2024 National Assessment of Educational Progress, 24% of fourth graders scored below *NAEP Basic* in mathematics, five percentage points more than in 2019 (National Center for Education Statistics, 2025). Learners in one class or one family also differ widely in what they are ready to learn next. National norms imply that one standard deviation of mathematics achievement within a grade spans about one year of typical growth in grade 3 and about *three to four years* in grade 8 (Bloom et al., 2008; Thum and Kuhfeld, 2020); AMP uses these figures for its own placement prior [S1]. One pace of instruction cannot meet all of those learners where they are.

The program. AMP (Adaptive Math Pathways) is an adaptive mathematics program from Corpora Inc. for the web, phones and tablets. Families, microschools, private schools and public schools can use it. A learner answers short batches of generated problems. After each answer AMP shows the result and a worked explanation, updates its estimate of the learner's working level in each strand of mathematics, and chooses the next problems. Assessment is therefore continuous: it is part of practice, not a separate test. Written lessons with worked examples and definitions, hints, narrated lesson films and printable workbooks draw on the same problem families. Points and a record of skills earned reward practice, and AMP is adding tools with which parents and teachers can set incentives of their own. Parents and teachers see each learner's record. The skill graph runs from

counting through precalculus [S2] and inventories all 229 numbered K–8 Common Core content standards [S3]. The system is described in [TN-1] and its content in [TN-3].

Population and setting. The rationale applies to learners at every grade AMP serves who practise with it regularly, at home or at school, alongside the mathematics instruction they already receive. In a school AMP supplements core instruction; in a home it may be the main practice program. Each study chooses its grade band and a normed test that measures it, and its analysis is stratified by grade band. Most of the planned measurement is in grades 3–8, where normed interim tests such as MAP Growth are most widely given.

3 Logic model

Figure 1 is the logic model. It reads from left to right: what AMP needs, what learners and adults do with it, what that activity produces that can be counted, and the changes in learning it is designed to cause. Each arrow is a claim that Section 4 supports with research and Section 7 plans to test. The assumptions and external factors along the bottom are the conditions under which the arrows are expected to hold. Section 6 says how each is monitored.

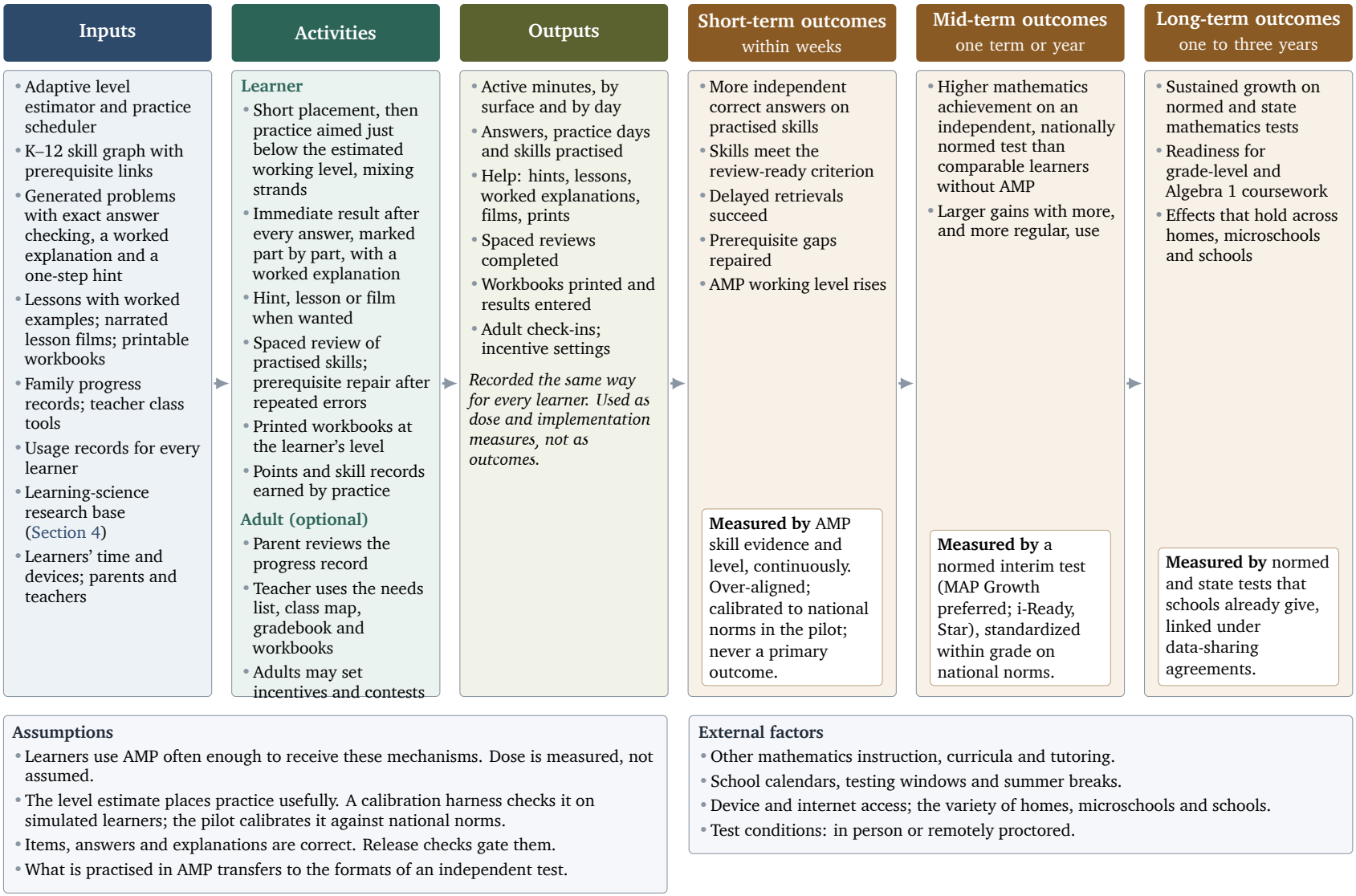


Figure 1. AMP logic model. Arrows are hypothesized causal links; Section 4 gives the research behind each and Section 5 the measure for each outcome.

4 Theory of change

AMP combines mechanisms that have been studied separately. The theory of change is that applying them together, to each learner's own evidence, raises both the number of productive practice opportunities a learner receives in an hour and the share of them that last. Table 1 lists the key components in the sense of 34 CFR 77.1, where AMP implements each, and the research behind it. The subsections give the detail and the limits.

4.1 C1. Practice at the learner's frontier

What AMP does. A new learner starts with four short scripted openers drawn from the core material for the grade expected at that age [S4]. From then on AMP keeps a Bayesian estimate of the learner's working level, with its uncertainty, on a curriculum-grade scale for each strand [S5, S6]. The scheduler aims new practice at the estimated level minus half a grade plus 0.3 times its uncertainty [S7, S8]. It therefore offers work the learner is likely, though not certain, to solve, and it reaches further when it knows less. Batches hold at most five problems, and the estimate is updated after every answer. The learner model is described in [TN-2].

Research. Vygotsky described the zone of proximal development as the distance between what a learner can do alone and what the learner can do with guidance (Vygotsky, 1978). Bjork and Bjork argue that conditions which make practice harder but still successful, which they call desirable difficulties, improve long-term retention (Bjork and Bjork, 2011). For a broad class of learning algorithms, Wilson et al. (2019) derive that "the optimal training accuracy is about 85%". That result is a theoretical model, not a classroom finding, but it formalizes the same principle: practice that is neither trivially easy nor mostly failed carries the most information for learning.

Link in the model. Targeting raises the share of practice time spent on problems the learner can learn from. It acts on every short-term outcome.

Limits. AMP's success targets are editorial parameters. The project's own documentation states that the estimator and scheduler "organize practice; they are not validated grade placement, mastery or standardized-test score estimates" [S9]. Calibration against real learners and national norms is planned (Section 5.1).

4.2 C2. Immediate, elaborated feedback

What AMP does. After every answer the server returns whether it was correct, the correct answer and a worked explanation of the method. When a learner chooses an authored wrong option, the explanation opens with feedback written for that choice [S10, S11]. Multi-part answers are marked blank by blank, and step-by-step work row by row: a derivation is checked for valid moves, so any valid path is accepted, and a proof is checked line by line against the reason given for each line [S12, S13, S14].

Table 1. Key components, their implementation and their research base.

Component	What AMP does	Research
C1 Practice at the learner's frontier	Estimates a working level per strand and aims practice just below it, reaching further when uncertain	Vygotsky (1978); Bjork and Bjork (2011); Wilson et al. (2019)
C2 Immediate, elaborated feedback	Result, correct answer and worked explanation after every answer; multi-part and step-by-step work marked part by part	Hattie and Timperley (2007); Shute (2008); Kluger and DeNisi (1996)
C3 Lessons and worked examples	A written lesson for each problem type: the method, its terms defined, and a worked example; every problem has a worked explanation; workbooks pair a lesson with guided problems	Sweller and Cooper (1985); Atkinson et al. (2000); Barbieri et al. (2023)
C4 Scaffolded hints	A one-step hint for every generated problem that names the first move and never the answer; hinted answers earn no independent credit	Wood et al. (1976); Aleven et al. (2003)
C5 Mastery criterion and prerequisite repair	A skill is review-ready only after independent success; repeated errors reserve work on prerequisites	Bloom (1968, 1984); Kulik et al. (1990)
C6 Retrieval, spacing and interleaving	Expanding review schedule from six hours to thirty days; batches mix strands	Roediger and Karpicke (2006); Adesope et al. (2017); Cepeda et al. (2006); Dunlosky et al. (2013); Rohrer and Taylor (2007)
C7 Narrated lesson films	Animated, narrated worked examples beside the lesson for each problem type	Mayer (2020); Höffler and Leutner (2007); Noetel et al. (2021); Guo et al. (2014)
C8 Progress visible to adults	Parent progress record; teacher needs list, class map and gradebook	Fuchs and Fuchs (1986); Black and Wiliam (1998); Kraft and Rogers (2015); Bergman (2021)
C9 Incentives and competition	Points for independent success, a record of skills earned, short animations; incentives and short contests that adults configure, at each learner's own level	Sailer and Homner (2020); Deci et al. (1999); Cameron et al. (2001); Levitt et al. (2016); Murayama and Elliot (2012)
All together: adaptive tutoring	The components above, in one system	Kulik and Fletcher (2016); Ma et al. (2014); VanLehn (2011); Steenbergen-Hu and Cooper (2013)

Research. In Hattie and Timperley's synthesis of meta-analyses on feedback, "the average effect size was 0.79" (Hattie and Timperley, 2007). Shute's review of formative feedback recommends feedback that is specific, focused on the task and elaborated with an explanation (Shute, 2008). Kluger and DeNisi found that feedback interventions "improved performance on average (d = .41) but that over 1/3 of the FIs decreased performance" (Kluger and DeNisi, 1996). Their analysis attributes the harm largely to feedback that directs attention to the self rather than the task. AMP's feedback is about the problem and its method only.

Link in the model. An error corrected immediately is not rehearsed. Feedback supports more independent correct answers and later retrieval of the right method.

4.3 C3. Lessons and worked examples

What AMP does. Each problem type has a written lesson. It states the goal, gives the plain-language meaning, defines the terms the method uses, sets out the steps, names the most common error, and presents a worked example read from question to answer [S15, S16, S17]. Words in lessons, hints and explanations open their glossary definitions [S18]. Every generated problem carries a worked explanation that the learner sees after answering (C2). A printed workbook pairs a lesson with guided problems from the same lesson, alongside practice and review chosen from the learner's own evidence [S19].

Research. Sweller and Cooper (1985) showed that algebra learners who studied worked examples learned faster than those who solved the equivalent problems. Atkinson et al. (2000) reviewed the instructional principles that make examples effective. In a meta-analysis of the worked-example effect on mathematics, Barbieri et al. (2023) report that "the average effect size of worked examples on mathematics performance outcomes was medium with g = 0.48".

Link in the model. Examples give a learner a correct method to imitate before and after an attempt, which supports independent success and repair of prerequisites.

4.4 C4. Scaffolded hints

What AMP does. Every generated problem has one teaching step written from its own givens. It names the problem's quantities and the first move of the method and never states the answer, a property a test checks for every problem type [S20, S21]. Where the problem has a figure, the hint can highlight the relevant parts of it. A learner who opens a hint or a lesson before answering has an *assisted* answer. It is recorded but earns no independent credit toward the mastery criterion (C5), and under the current default policy it carries no weight in the level estimate [S22, S6]. Each hint, lesson, worked explanation, film and print opened is recorded as help [S23].

Research. Wood, Bruner and Ross introduced *scaffolding*: a tutor controls the parts of a task beyond the learner's current capacity so that the learner can complete the rest (Wood et al., 1976). A review of help seeking in interactive learning environments found that "learners are not using help facilities effectively" (Aleven et al., 2003). AMP therefore keeps hints to a single step, withholds independent

credit for helped answers so that help cannot replace learning, and records the use of help so that its effect can be studied.

Link in the model. A hint turns a problem slightly beyond reach into one the learner can complete, extending the range of productive practice (C1).

4.5 C5. Mastery criterion and prerequisite repair

What AMP does. A skill is *review-ready* only after at least five attempts, with four of the latest five independently correct and the latest independently correct [S22]. Errors and assisted answers return it to *practising*. When a learner misses a skill twice in the recent window, the scheduler reserves work on its prerequisites, following the skill graph to the nearest skill with problems available [S24, S8]. Success on a skill never grants credit for prerequisites that were not observed.

Research. Bloom (1968) argued that "most students, perhaps over 90 percent, can master what teachers have to teach them" when time and corrective help vary with need. Bloom (1984) framed one-to-one tutoring with mastery methods as the benchmark that group instruction should try to approach. Kulik, Kulik and Bangert-Drowns's meta-analysis of "108 controlled evaluations showed that mastery learning programs have positive effects on the examination performance" of students (Kulik et al., 1990).

Link in the model. Moving on only after independent success, and stepping back after repeated errors, keeps practice built on secure foundations. This mechanism produces the short-term outcomes "skills meet the review-ready criterion" and "prerequisite gaps repaired".

Limits. Classic mastery learning holds a group in fixed units with formative tests and correctives. AMP applies a mastery criterion to individual skills and does not hold a learner back in other strands. Its thresholds are editorial and have not yet been evaluated against learners.

4.6 C6. Retrieval practice, spacing and interleaving

What AMP does. Each practised skill has a review schedule with intervals of six hours, one day, three days, seven days, fourteen days and thirty days [S25]. Independent success advances the schedule; errors and assistance reset it. Due reviews reserve places in practice batches. The scheduler penalizes repeating a skill or strand within a batch, so practice interleaves strands rather than blocking one topic [S8]. Printed workbooks include a review section of due skills [S19].

Research. Retrieval practice: Roediger and Karpicke (2006) found that "on the delayed tests, prior testing produced substantially greater retention than studying". A meta-analysis of practice testing concluded that "practice tests are more beneficial for learning than restudying and all other comparison conditions" (Adesope et al., 2017). Spacing: a quantitative synthesis of 839 assessments in 317 experiments found that the gap between study sessions that maximizes retention grows as the retention interval grows (Cepeda et al., 2006), which is the reasoning behind an expanding schedule. Dunlosky et al. (2013) reviewed ten learning techniques and rated practice testing and distributed practice as the two of highest utility. Interleaving: in Rohrer and Taylor (2007), learn-

ers who practised mixed mathematics problems outperformed those who practised in blocks on a delayed test (63% against 20%, $d = 1.34$).

Link in the model. Spaced, mixed retrieval converts initial success into durable, discriminating knowledge: the learner must recall which method applies, not only how to execute it. This mechanism produces "delayed retrievals succeed" and underlies transfer to a mixed-content normed test.

4.7 C7. Narrated lesson films

What AMP does. AMP's lesson films are animated, narrated worked examples. The figure is drawn in step with the narration [S26]. A film sits beside the written lesson for its problem type and can be opened from practice and from the lesson library [S27, S28]. At the commit cited here the library holds 154 films serving 512 problem types [S29], and it grows every week. Opening a film, playing it to the end and the time spent watching are recorded [S28, S30].

Research. Mayer's cognitive theory of multimedia learning explains how narration timed to the picture it describes supports learning (Mayer, 2020). In a meta-analysis of 76 comparisons, Höffler and Leutner (2007) found an advantage of instructional animations over static pictures ($d = 0.37$). A systematic review in higher education found that replacing other teaching methods with video gave a small benefit ($g = 0.28$) and adding video to existing teaching a larger one ($g = 0.80$) (Noetel et al., 2021). In an observational study of online course videos, Guo et al. (2014) found that shorter videos held learners' engagement better.

Link in the model. A film is a worked example (C3) in motion. It turns a problem the learner cannot yet start into one whose method they have just seen worked.

Limits. The experimental evidence comes mostly from older learners. AMP's films are new, and their own effect has not yet been estimated.

4.8 C8. Progress visible to parents and teachers

What AMP does. Parents see each child's progress record: levels, skills, time by day and skill, and workbooks [S31]. Teachers have a daily list of who needs them and which students share a gap, a class map, a gradebook in which every grade opens to its evidence, and a record of the help each student used [S32]. Adults can print a workbook at the learner's level and enter its results, which feed the same evidence [S33]. Each adult check-in is recorded [S34].

Research. In a meta-analysis of systematic formative evaluation, in which teachers monitor progress data and adjust instruction, Fuchs and Fuchs (1986) report "an average weighted effect size of .70". Black and Wiliam's review concluded that strengthening frequent feedback in classrooms yields substantial learning gains (Black and Wiliam, 1998). In a field experiment, frequent teacher-to-parent messages "decreased the percentage of students who failed to earn course credit from 15.8% to 9.3%" (Kraft and Rogers, 2015). Bergman (2021) found that giving parents information about their children's academic progress "improves student achievement".

Link in the model. Visible progress lets an adult notice a stalled learner, encourage regular practice and teach to a shared gap. Adult involvement is optional: many learners use AMP on their own, and the other components act without it.

4.9 C9. Incentives and competition

What AMP does. A correct answer without help earns a point, shown with the result, and fills the bar for that skill in the learner's record of skills [S35, S36]. Short animations mark a correct result and a mastered skill [S37]. AMP is adding tools with which a parent or teacher can set incentives of their own, such as breaks earned by practice or short contests within a class session. Because every learner practises at their own level, a contest can reward effort and accuracy at that level rather than prior achievement. Adults decide whether and how to use these tools, and the settings they choose are recorded with use.

Research. In a meta-analysis of gamification in learning, Sailer and Homner (2020) report effects on cognitive (g = .49), motivational (g = .36) and behavioural (g = .25) outcomes. Rewards also carry a known risk. In a meta-analysis of experiments, expected tangible rewards for doing a task lowered later free-choice interest, more so in children, while positive feedback raised it (Deci et al., 1999); rewards tied to a level of performance did not lower it (Cameron et al., 2001). Across four decades of studies, incentives predicted the quantity of performance and intrinsic motivation its quality (Cerasoli et al., 2014). Timing matters: in field experiments with school students, rewards given immediately raised test scores by about a tenth of a standard deviation, and the same rewards delayed by a month had no effect (Levitt et al., 2016). Competition helps some learners and hurts others. Its average relation to performance is near zero (Murayama and Elliot, 2012), and a public leaderboard introduced into high-school online courses was followed by lower performance, most of all among prior top performers (Bursztyn and Jensen, 2015). Tournaments with rules tilted toward less-prepared children did not lower either group's effort and spread the wins (Calsamiglia et al., 2013).

Link in the model. Incentives act on the amount and regularity of use, the dose through which the other components work (Section 4.12). The research suggests that their effect depends on design: immediate, tied to effort or accuracy at the learner's level, and short in the case of competition. The best settings are not yet known. Adults will configure incentives in different ways; those choices are recorded, and their relation to use and growth generates hypotheses for later randomized tests (Section 7).

4.10 The combination: evidence on adaptive tutoring systems

AMP's components together form an adaptive tutoring system, and that class of programs has its own evidence. Kulik and Fletcher (2016) found that the median effect of intelligent tutoring systems across 50 controlled evaluations "was to raise test scores 0.66 standard deviations over conventional levels". Ma et al. (2014) report advantages over teacher-led, large-group instruction (g = .42), non-adaptive computer-based instruction (g = .57) and textbooks or workbooks (g = .35). VanLehn (2011) found that "the effect size of intelligent tutoring systems was 0.76, so they are nearly as effective as human tutoring" (0.79). VanLehn classifies systems by the grain of interaction: answer-

based, step-based and substep-based. AMP spans the first two. Many of its problems take one answer; its multi-part and step-by-step problems are marked part by part and step by step (C2).

Kulik and Fletcher also report that the size of the improvement "depended to a great extent on whether improvement was measured on locally developed or standardized tests". That is why AMP's planned evaluations use an independent standardized test as the primary outcome. On such tests the meta-analytic record for intelligent tutoring in K–12 mathematics is mixed, with "average effect sizes ranging from g = 0.01 to g = 0.09" (Steenbergen-Hu and Cooper, 2013), while strong programs in large field trials have shown substantial gains: Cognitive Tutor Algebra I raised the median student's performance by about eight percentile points in its second year (Pane et al., 2014), and online mathematics homework with immediate feedback produced $g = 0.18$ on a nationally normed test (Roschelle et al., 2016). Against benchmarks drawn from randomized trials with achievement outcomes, an effect of 0.20 standard deviations or more counts as large (Kraft, 2020); see also Hill et al. (2008). The rationale therefore predicts a positive effect on an independent normed test, growing with use.

4.11 How the components combine

The components act on the same unit, one practice opportunity. Targeting (C1) chooses a problem the learner can probably solve with effort. A hint (C4), a worked example (C3) or a film (C7) makes it solvable when it is not. Feedback (C2) corrects an error before it is rehearsed. Spaced review and interleaving (C6) bring the skill back after a delay and among other kinds of problems. The mastery criterion and repair (C5) decide whether to move on or step back. Visibility (C8) lets an adult see that sequence and act. Incentives (C9) bring the learner back to the next opportunity. Because AMP records each step, the evaluation can relate the dose of each component to outcomes as well as estimate the effect of the whole program.

4.12 Dose

Each component acts only while a learner is practising, so the theory predicts that gains grow with use. Carroll's model of school learning makes the degree of learning a function of the time a learner actually spends relative to the time the learner needs (Carroll, 1963). How use is spread may matter as much as how much there is: the same practice time spread across more sessions is retained better (Cepeda et al., 2006). Short, regular sessions may therefore do more than fewer, longer ones, and there may be a point past which more minutes in a day add little. AMP records both the amount and the pattern of each learner's use: active hours, active days a week, minutes per day and the gaps between days [TN-4]. The evaluation relates growth to each, and the pilot study compares practice schedules directly in a randomized experiment [RR-3].

5 Outcomes and how each will be measured

Each outcome in the logic model has a measure and a source. The primary outcome of every planned impact study is an independent, nationally normed test that AMP did not write. AMP's own measures serve for dose, implementation and proximal learning, and are labelled over-aligned wherever they appear.

Table 2. Outcomes, measures and their role in the evaluation.

Logic-model element	Measure and source	When	Role
Outputs: use and help	Active minutes by surface and by day; answers; help, including films; workbooks; adult check-ins [S38, S23, S34]. Defined in [TN-4]	Continuous	Dose and implementation fidelity
Short-term: independent accuracy, review-ready skills, delayed retrievals	Skill-evidence summaries: attempts, independent correct answers, latest five outcomes, retrieval schedule [S39, S22]	Continuous	Proximal; over-aligned
Short-term: AMP working level	Stored level and growth over a window, with its standard error [S40]	Continuous; at each normed test	Calibrated to national norms in the pilot; secondary outcome in the trials, measured in both arms
Mid-term: mathematics achievement	A normed interim test: MAP Growth Mathematics (preferred), i-Ready Diagnostic or Star Math, standardized within grade on each test's national norms (Thum and Kuhfeld, 2020)	Start and end of the study period	Primary outcome
Mid-term: gains grow with use	Standardized growth against active hours and the pattern of use	Start and end of the study period	Confirmatory in the pilot; second hypothesis in the trials
Long-term: sustained achievement	Normed interim tests that schools already give and state assessments, linked under data-sharing agreements	Each year	Outcome of the follow-up and comparison-group studies
Long-term: course readiness	Course placement or enrolment records, where a partner can share them	One to three years	Exploratory; to be specified in a later protocol

5.1 Continuous assessment and calibration

AMP assesses as it teaches. Every answer updates the level estimate, so a learner's position and growth are known at any time without a separate test session [S40]. That record is aligned to AMP's own practice, and its parameters are editorial, so it is not yet tied to any national scale [S9]. Calibration is planned as the first step of the evaluation. The pilot study links AMP's level, measured within two weeks of each normed test, to MAP Growth scores and national percentiles in learners who have both, and publishes the crosswalk with its prediction intervals [RR-3]. AMP's continuous record can then be read on a national scale between normed tests. It remains a secondary measure; the primary outcome of every impact study is the independent test.

6 Assumptions and external factors

Each assumption could fail. Table 3 states how each will be monitored so that a null result can be interpreted.

Table 3. Assumptions and how each is monitored.

Assumption	How it is monitored
Learners use AMP enough to receive the mechanisms	Active minutes, active days and answers are recorded for every learner, the same way in every setting [S38, S41, S42]. The trials estimate the effect of the offer of AMP and the effect of its use.
The level estimate places practice usefully	A calibration harness runs simulated learners through the production estimator and reports bias, error and interval coverage [S43]. The pilot study relates AMP's level to national norms in real learners (Section 5.1).
Items, answers and explanations are correct	Exact solutions and per-format answer checkers [S44]; a test that no hint gives away its answer [S21]; instructional review before release [TN-3]. A learner or an adult can flag a problem [S45].
Practice in AMP transfers to independent tests	Tested directly: the primary outcome is a normed test AMP did not write. AMP's own level is reported separately and labelled over-aligned.

External factors. Other mathematics instruction, curricula and tutoring differ across learners. In the randomized studies they are balanced in expectation by assignment and described at baseline. In the non-randomized studies they are a source of confounding that the design addresses with baseline scores and prior achievement as controls and with matched comparison groups [RR-3, RR-4]. School calendars, testing windows and summer breaks constrain when studies can run. Device and internet access is a condition of use. Settings range from single families to schools, so setting is recorded and enters each analysis. Test conditions differ between in-person and remotely proctored administration, so each study records the mode and, where there are two groups, uses the same mode for both.

7 Planned evaluation

The statute requires "ongoing efforts to examine the effects" of a Tier 4 activity, and the 2016 guidance asks for an effort "ideally producing promising evidence or higher". Corpora plans a phased program of studies that climb the evidence levels and share one data infrastructure [RR-1]. Each is pre-registered as its own study, and each gives the next its design inputs.

Table 4. The phases of the evaluation.

Phase	When	What	Evidence level sought
Rationale (this document)	Now	Logic model; use recorded equally for every learner	Tier 4, Demonstrates a Rationale
Pilot study [RR-3]	Rolling enrollment; confirmatory analysis in 2027	Calibration against national norms; growth with use, with statistical controls; embedded randomized experiments	Tier 3, Promising
Comparison groups [RR-4]	2027–28 school year, or earlier from existing school data	AMP users against matched non-users in schools that already test	Tier 2, Moderate
Randomized trials [RR-5]	Fall 2027 at the earliest	Family and school modules under one master protocol	Tier 1, Strong

7.1 Pilot study (toward Tier 3)

The pilot study is specified in [RR-3]. Learners in any grade AMP covers, at home or in school, receive AMP for the study period and take a normed mathematics test at the start and the end. The study first calibrates AMP's level against national norms (Section 5.1). Its confirmatory hypothesis is that standardized growth rises with active hours of use, controlling for baseline score, grade, test, time between tests and setting, as the definition of promising evidence requires ("a correlational study with statistical controls for selection bias") (U.S. Department of Education, 2024). Randomized experiments are embedded at no cost to participants: a lottery for start dates when applications exceed places, randomized supports for regular use, and component experiments inside AMP, among them practice schedules (Section 4.12).

The incentive settings that parents and teachers choose are recorded, and their relation to use and growth is analysed as exploratory, to generate hypotheses for later randomized tests.

7.2 Comparison-group studies (toward Tier 2)

In schools that already give a normed interim test, AMP users will be compared with matched non-users in the same school or district, prospectively or from the previous year's data, with baseline equivalence shown on the prior mathematics score under the What Works Clearinghouse standards (What Works Clearinghouse, 2022). Each condition will include several teachers and schools so that no single classroom is confounded with the program. Tier 2 requires at least 350 students across more than one site (U.S. Department of Education, 2024). The design is specified in [RR-4].

7.3 Randomized trials (toward Tier 1)

A master protocol governs the randomized trials [RR-5]. Its modules are a family-randomized waitlist and a school- or classroom-randomized delayed start, with each site's parameters set in an addendum registered before randomization.

- **Participants.** Learners in any grade with a matching normed test; randomization is blocked by grade band and setting, and siblings stay together.
- **Outcomes.** Primary: a normed mathematics test, standardized within grade. Secondary: AMP's level in both arms, labelled over-aligned.
- **Hypotheses.** First, that AMP raises mathematics achievement, reported as the effect of the offer (intention to treat) and the effect of use. Second, tested only if the first is significant, that the effect grows with use.
- **Size.** Tier 1 requires a study that meets the What Works Clearinghouse standards without reservations (What Works Clearinghouse, 2022), with at least 350 students across more than one site (U.S. Department of Education, 2024). Sample sizes are set by the power analysis in [RR-5].
- **Credibility.** Pre-registration, an independent analyst, and publication of results whatever they show.

7.4 Implementation monitoring

Throughout every study the outputs in Figure 1 are recorded continuously. They show whether the program was delivered as designed and let the evaluation separate a weak program from weak implementation.

7.5 How this document meets the Tier 4 criteria

Table 5. Tier 4 criteria and where they are met. Section A quotes each criterion in full.

Criterion and source	Where it is met
A logic model that identifies key project components and their relationships to relevant outcomes (34 CFR 77.1)	Figure 1; components C1–C9 in Table 1; Section 4
Key components supported by citations of high-quality research or evaluation findings suggesting they are likely to significantly improve relevant outcomes (34 CFR 77.1, 2024)	Section 4: peer-reviewed experiments and meta-analyses for each component, with effect sizes quoted where the source reports one
Rationale likely to improve student outcomes (ESEA 8101(21)(A)(ii)(I))	Section 4; relevant outcomes defined in Section 5
Ongoing efforts to examine the effects (ESEA 8101(21)(A)(ii)(II))	Section 7: pilot study, comparison-group studies and randomized trials; continuous implementation monitoring

Criterion and source	Where it is met
A well-specified logic model, and an effort to study effects ideally producing promising evidence or higher (2016 guidance)	Figure 1; Sections 5 and 7; the planned studies target Tiers 3, 2 and 1
Logic-model elements: resources, activities, outputs, short-, mid- and long-term outcomes, assumptions (REL toolkits)	Figure 1; Section 6

8 Limits of this rationale

- No study has yet estimated AMP's effect on student outcomes. This document establishes a rationale; the studies in Section 7 estimate effectiveness.
- The research cited supports each component in other programs and populations. Effects of separate components do not simply add up.
- AMP's targeting parameters, mastery thresholds and review intervals are editorial choices. The pilot study is the first step in fitting them to real learners.
- AMP's own level is aligned to its own practice. It is never used as the primary outcome.
- Parents and teachers use AMP in different ways and configure its incentives differently. How adults implement AMP may matter as much as the program, and it is measured, not controlled.
- The logic model will be revised after each study. Results will be published whether they are positive, null or negative.

References

Olusola O. Adesope, Dominic A. Trevisan, and Narayankripa Sundararajan. Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3):659–701, 2017. doi: 10.3102/0034654316689306.

Vincent Aleven, Elmar Stahl, Silke Schworm, Frank Fischer, and Raven Wallace. Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73(3):277–320, 2003. doi: 10.3102/00346543073003277.

Robert K. Atkinson, Sharon J. Derry, Alexander Renkl, and Donald Wortham. Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70(2):181–214, 2000. doi: 10.3102/00346543070002181.

Christina Areizaga Barbieri, Dana Miller-Cotto, Sarah N. Clerjuste, and Kamal Chawla. A meta-analysis of the worked examples effect on mathematics performance. *Educational Psychology Review*, 35(1):11, 2023. doi: 10.1007/s10648-023-09745-1.

Peter Bergman. Parent-child information frictions and human capital investment: Evidence from a field experiment. *Journal of Political Economy*, 129(1):286–322, 2021. doi: 10.1086/711410.

Elizabeth Ligon Bjork and Robert A. Bjork. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, pages 56–64. Worth Publishers, New York, 2011. URL https://bjorklab.psych.ucla.edu/wp-content/uploads/sites/13/2016/04/EBjork_RBjork_2011.pdf.

Paul Black and Dylan Wiliam. Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1):7–74, 1998. doi: 10.1080/0969595980050102.

Benjamin S. Bloom. Learning for mastery. *Evaluation Comment*, 1(2), May 1968. URL https://eric.ed.gov/?id=ED053419. University of California, Los Angeles, Center for the Study of Evaluation of Instructional Programs. ERIC ED053419.

Benjamin S. Bloom. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6):4–16, 1984. doi: 10.3102/0013189X013006004.

Howard S. Bloom, Carolyn J. Hill, Alison Rebeck Black, and Mark W. Lipsey. Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1(4):289–328, 2008. doi: 10.1080/19345740802400072.

Leonardo Bursztyn and Robert Jensen. How does peer pressure affect educational investments? *Quarterly Journal of Economics*, 130(3):1329–1367, 2015. doi: 10.1093/qje/qjv021.

Caterina Calsamiglia, Jörg Franke, and Pedro Rey-Biel. The incentive effects of affirmative action in a real-effort tournament. *Journal of Public Economics*, 98:15–31, 2013. doi: 10.1016/j.jpubeco.2012.11.003.

Judy Cameron, Katherine M. Banko, and W. David Pierce. Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst*, 24(1):1–44, 2001. doi: 10.1007/BF03392017.

John B. Carroll. A model of school learning. *Teachers College Record*, 64(8), 1963. doi: 10.1177/016146816306400801.

Nicholas J. Cepeda, Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3):354–380, 2006. doi: 10.1037/0033-2909.132.3.354.

Christopher P. Cerasoli, Jessica M. Nicklin, and Michael T. Ford. Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, 140(4):980–1008, 2014. doi: 10.1037/a0035661.

Edward L. Deci, Richard Koestner, and Richard M. Ryan. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6):627–668, 1999. doi: 10.1037/0033-2909.125.6.627.

John Dunlosky, Katherine A. Rawson, Elizabeth J. Marsh, Mitchell J. Nathan, and Daniel T. Willingham. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1):4–58, 2013. doi: 10.1177/1529100612453266.

Lynn S. Fuchs and Douglas Fuchs. Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3):199–208, 1986. doi: 10.1177/001440298605300301.

Philip J. Guo, Juho Kim, and Rob Rubin. How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 41–50. ACM, 2014. doi: 10.1145/2556325.2566239.

John Hattie and Helen Timperley. The power of feedback. *Review of Educational Research*, 77(1): 81–112, 2007. doi: 10.3102/003465430298487.

Carolyn J. Hill, Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3):172–177, 2008. doi: 10.1111/j.1750-8606.2008.00061.x.

Tim N. Höffler and Detlev Leutner. Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17(6):722–738, 2007. doi: 10.1016/j.learninstruc.2007.09.013.

Wendy Kekahio, Brian Lawton, Louis Cicchinelli, and Paul R. Brandon. Logic models: A tool for effective program planning, collaboration, and monitoring. REL 2014–025. Regional Educational Laboratory Pacific, Institute of Education Sciences, 2014. URL https://ies.ed.gov/sites/default/files/migrated/rel/regions/pacific/resources/elm/pdf/LogicModelsELM_effectiveProgramPlanning.pdf.

Avraham N. Kluger and Angelo DeNisi. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2):254–284, 1996. doi: 10.1037/0033-2909.119.2.254.

Matthew A. Kraft. Interpreting effect sizes of education interventions. *Educational Researcher*, 49 (4):241–253, 2020. doi: 10.3102/0013189X20912798.

Matthew A. Kraft and Todd Rogers. The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review*, 47:49–63, 2015. doi: 10.1016/j.econedurev.2015.04.001.

Chen-Lin C. Kulik, James A. Kulik, and Robert L. Bangert-Drowns. Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60(2):265–299, 1990. doi: 10.3102/00346543060002265.

James A. Kulik and J. D. Fletcher. Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1):42–78, 2016. doi: 10.3102/0034654315581420.

Brian Lawton, Paul R. Brandon, Louis Cicchinelli, and Wendy Kekahio. Logic models: A tool for designing and monitoring program evaluations. Technical Report REL 2014–007, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Pacific, Washington, DC, February 2014. URL https://eric.ed.gov/?id=ED544752.

Steven D. Levitt, John A. List, Susanne Neckermann, and Sally Sadoff. The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, 8(4):183–219, 2016. doi: 10.1257/pol.20130358.

Wenting Ma, Olusola O. Adesope, John C. Nesbit, and Qing Liu. Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4):901–918, 2014. doi: 10.1037/a0037123.

Richard E. Mayer. *Multimedia Learning*. Cambridge University Press, Cambridge, 3 edition, 2020. ISBN 9781107187504. doi: 10.1017/9781316941355.

Kou Murayama and Andrew J. Elliot. The competition–performance relation: A meta-analytic review and test of the opposing processes model of competition and performance. *Psychological Bulletin*, 138(6):1035–1070, 2012. doi: 10.1037/a0028324.

National Center for Education Statistics. The nation's report card: 2024 mathematics assessment, grades 4 and 8. U.S. Department of Education, Institute of Education Sciences, 2025. URL https://www.nationsreportcard.gov/reports/mathematics/2024/g4_8/. Accessed 2026-11-17.

Michael Noetel, Shantell Griffith, Oscar Delaney, Taren Sanders, Philip Parker, Borja del Pozo Cruz, and Chris Lonsdale. Video improves learning in higher education: A systematic review. *Review of Educational Research*, 91(2):204–236, 2021. doi: 10.3102/0034654321990713.

John F. Pane, Beth Ann Griffin, Daniel F. McCaffrey, and Rita Karam. Effectiveness of Cognitive Tutor Algebra I at scale. *Educational Evaluation and Policy Analysis*, 36(2):127–144, 2014. doi: 10.3102/0162373713507480.

Regional Educational Laboratory Midwest. ESSA tiers of evidence: What you need to know. Handout, American Institutes for Research, contract ED-IES-17-C-0007, 2019. URL https://web.archive.org/web/20260314062158id_/https://www.azed.gov/sites/default/files/2021/01/Handout%201%20REL%20Midwest%20ESSA%20Tiers%20of%20Evidence.pdf.

Henry L. Roediger and Jeffrey D. Karpicke. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3):249–255, 2006. doi: 10.1111/j.1467-9280.2006.01693.x.

Doug Rohrer and Kelli Taylor. The shuffling of mathematics problems improves learning. *Instructional Science*, 35(6):481–498, 2007. doi: 10.1007/s11251-007-9015-8.

Jeremy Roschelle, Mingyu Feng, Robert F. Murphy, and Craig A. Mason. Online mathematics homework increases student achievement. *AERA Open*, 2(4):2332858416673968, 2016. doi: 10.1177/2332858416673968.

Michael Sailer and Lisa Homner. The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1):77–112, 2020. doi: 10.1007/s10648-019-09498-w.

Karen Shakman and Sheila M. Rodriguez. Logic models for program design, implementation, and evaluation: Workshop toolkit. Technical Report REL 2015–057, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands, Washington, DC, May 2015. URL https://eric.ed.gov/?id=ED556231.

Valerie J. Shute. Focus on formative feedback. *Review of Educational Research*, 78(1):153–189, 2008. doi: 10.3102/0034654307313795.

Saiying Steenbergen-Hu and Harris Cooper. A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, 105 (4):970–987, 2013. doi: 10.1037/a0032447.

John Sweller and Graham A. Cooper. The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2(1):59–89, 1985. doi: 10.1207/s1532690xci0201_3.

Yeow Meng Thum and Megan Kuhfeld. NWEA 2020 MAP Growth achievement status and growth norms for students and schools. NWEA Research Report, NWEA, Portland, OR, 2020. URL https://teach.mapnwea.org/impl/normsResearchStudy.pdf.

U.S. Code. 20 U.S.C. 7801(21): Definitions, "Evidence-based" (ESEA Section 8101(21), as Amended by the Every Student Succeeds Act), 2015. URL https://www.law.cornell.edu/uscode/text/20/7801.

U.S. Department of Education. Non-regulatory guidance: Using evidence to strengthen education investments. September 16, 2016; archived, superseded by the 2023 revision, 2016. URL https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseseinvestment.pdf.

U.S. Department of Education. Non-regulatory guidance: Using evidence to strengthen education investments. Office of Planning, Evaluation and Policy Development. Revised September 28, 2023, 2023. URL https://www.ed.gov/sites/ed/files/fund/grant/about/discretionary/2023-non-regulatory-guidance-evidence.pdf.

U.S. Department of Education. Education department general administrative regulations, 34 CFR 77.1: Definitions that apply to all department programs. Electronic Code of Federal Regulations, as amended by 89 FR 70340 (August 29, 2024), 2024. URL https://www.ecfr.gov/current/title-34/subtitle-A/part-77/section-77.1.

Kurt VanLehn. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4):197–221, 2011. doi: 10.1080/00461520.2011.611369.

L. S. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA, 1978. doi: 10.2307/j.ctvjf9vz4. URL https://www.hup.harvard.edu/books/9780674576292. Edited by M. Cole, V. John-Steiner, S. Scribner and E. Souberman.

What Works Clearinghouse. What works clearinghouse procedures and standards handbook, version 5.0. Technical report, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Washington, DC, 2022. URL https://ies.ed.gov/ncee/WWC/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf. Revised December 2022.

Robert C. Wilson, Amitai Shenhav, Mark Straccia, and Jonathan D. Cohen. The Eighty Five Percent Rule for optimal learning. *Nature Communications*, 10(1):4646, 2019. doi: 10.1038/s41467-019-12552-4.

David Wood, Jerome S. Bruner, and Gail Ross. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2):89–100, 1976. doi: 10.1111/j.1469-7610.1976.tb00381.x.

A Tier 4 checklist: each criterion and where it is met

Each federal criterion for ESSA's fourth evidence level, quoted in full. Page numbers refer to this document.

Statute: ESEA section 8101(21)(A)(ii), 20 U.S.C. 7801(21)

Criterion (verbatim)	How it is met	Where
"(I) demonstrates a rationale based on high-quality research findings or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes or other relevant outcomes"	Eight key components (C1–C8), each tied to peer-reviewed experiments or meta-analyses, with effect sizes where the source reports one; evidence on adaptive tutoring as a whole	Table 1, p. 6; Section 4, pp. 5–11
"(II) includes ongoing efforts to examine the effects of such activity, strategy, or intervention"	A phased program: a pilot study, comparison-group studies and randomized trials, with continuous implementation monitoring	Section 7, p. 13

Regulation: 34 CFR 77.1, as amended August 29, 2024

Criterion (verbatim)	How it is met	Where
Demonstrates a rationale: "a key project component included in the project's logic model that is supported by citations of high-quality research or evaluation findings that suggest that the project component is likely to significantly improve relevant outcomes"	Each component appears in the logic model and has its own research paragraph with citations	Figure 1, p. 4; Section 4
Logic model: "a framework that identifies key project components … and describes the theoretical and operational relationships among the key project components and relevant outcomes"	Theoretical relationships: the research for each component and how they combine. Operational relationships: what AMP does, cited to its source code at one commit	Figure 1; Section 4; Section B, p. 24
Relevant outcome: "the student outcome(s) or other outcome(s) the key project component is designed to improve"	Short-, mid- and long-term outcomes, each with a measure, timing and role; the primary outcome is an independent normed test	Table 2, p. 12
Project component: "an activity, strategy, intervention, process, product, practice, or policy … Evidence may pertain to an individual project component or to a combination"	Evidence for each component and for the combination	Section 4; Section 4.10, p. 10

Non-regulatory guidance, revised September 28, 2023

Criterion (verbatim)	How it is met	Where
Table 1, Demonstrates a Rationale: "Logic model informed by research or evaluation findings"	A logic model with a research base for each component	Figure 1; Section 4
p. 15: "a key project component included in the project's logic model is informed by research or evaluation findings that suggest the project component is likely to improve relevant outcomes"	As above	Table 1
p. 9: "A logic model, informed by high-quality evidence, that suggests how the project component is likely to improve relevant outcomes" and "SMART (Specific, Measurable, Achievable, Relevant, and Time-Bound) goals"	Outcomes with a named instrument and timing; phases with dates and evidence targets	Tables 2 and 4

Original guidance, September 16, 2016 (archived; still reproduced by many states)

Criterion (verbatim)	How it is met	Where
"1) A well-specified logic model that is informed by research or an evaluation that suggests how the intervention is likely to improve relevant outcomes"	All toolkit elements; outcomes tied to measures; assumptions with checks	Figure 1; Sections 4 to 6
"2) An effort to study the effects of the intervention, ideally producing promising evidence or higher, that will happen as part of the intervention or is underway elsewhere"	The phases target Tier 3 (Promising), Tier 2 (Moderate) and Tier 1 (Strong)	Section 7; Table 4, p. 14

Logic-model elements (Regional Educational Laboratory toolkits)

Element	Where
Problem statement (Shakman and Rodriguez, 2015, p. 16)	Section 2, p. 2
Resources (inputs)	Figure 1

Element	Where
Strategies and activities	Figure 1; Section 4
Outputs (activity data for tracking implementation)	Figure 1; Table 2
Short-, mid- and long-term outcomes	Figure 1; Table 2
Assumptions and external factors	Figure 1; Table 3, p. 13

B Where AMP implements each component

Each citation opens the file at the commit named on the title page.

Component or record	Source artifacts
Placement	[S4]
Level estimate (C1)	[S5, S6]
Targeting and batch selection (C1, C6)	[S7, S8]
Feedback and answer checking (C2)	[S12, S44, S13, S14, S10, S11]
Lessons and worked examples (C3)	[S15, S16, S17, S18]
Hints (C4)	[S20, S21]
Mastery criterion and repair (C5)	[S22, S24]
Spaced review (C6)	[S25, S39]
Lesson films (C7)	[S26, S27, S29, S28]
Parent and teacher views (C8)	[S31, S32, S34]
Incentives (C9)	[S35, S36, S37]
Workbooks	[S19, S33]
Use records	[S38, S41, S23, S42]
Calibration and norms	[S43, S1, S40]

Documents in this series cited here

[RR-1] *Building the evidence for AMP: a phased research program*, version 1.0. Corpora Inc. Distributed as `amp-rr1-research-program.pdf`.

[TN-1] *How AMP works: system description*, version 1.0. Corpora Inc. Distributed as `amp-tn1-system-description.pdf`.

[TN-2] *Learner model, placement and national norms*, version 1.0. Corpora Inc. Distributed as `amp-tn2-learner-model.pdf`.

[TN-3] *Curriculum, lessons, videos and workbooks*, version 1.0. Corpora Inc. Distributed as `amp-tn3-curriculum.pdf`.

[TN-4] *Measuring use: instrumentation and data dictionary*, version 1.0. Corpora Inc. Distributed as `amp-tn4-instrumentation.pdf`.

[TN-5] *Open research and development questions*, version 1.0. Corpora Inc. Distributed as `amp-tn5-rd-questions.pdf`.

[RR-3] *The AMP Pilot Study: use, growth and calibration against national norms. Protocol and analysis plan*, version 1.0. Corpora Inc. Distributed as `amp-rr3-pilot-study-protocol.pdf`.

[RR-4] *Comparison-group studies of AMP in schools. Protocol and analysis plan*, version 1.0. Corpora Inc. Distributed as `amp-rr4-comparison-group-protocol.pdf`.

[RR-5] *AMP and mathematics achievement: master protocol and pre-analysis plan for randomized trials*, version 2.0. Corpora Inc. Distributed as `amp-rr5-randomized-trial-protocol.pdf`.

Source artifacts

Statements about what AMP does are cited to its source code and documentation, which are public on GitHub at github.com/corpora-inc/encorpora. All citations refer to one commit, dated 2026-11-18, so they stay valid as the software changes:

`7fd3147f0092ed6d7d4fe0f9ace1cfca64f65e69`

Each path below links to that file at that commit; where a line range is given, the link opens at those lines.

	What it shows	File at the cited commit
[S1]	Within-grade spread of mathematics achievement in grade-years, from national norms	`amp/src/schools/norms-prior.ts` (line 11)
[S2]	Skill graph: objectives and prerequisite links	`amp/src/curriculum/catalog.ts` (line 1020)
[S3]	Curriculum inventory: all 229 numbered K–8 Common Core content standards	`amp/docs/curriculum.md` (line 17)
[S4]	Placement: four scripted openers, then adaptive selection; the first 20 answers are the assessment window	`amp/src/learning/placement.ts` (lines 31–35)
[S5]	Level estimate: a Bayesian working level, with its uncertainty, for each strand	`amp/src/learning/frontier.ts` (line 48)
[S6]	The production estimator's policy (`grade-frontier-v4-evidence-anchored-climb`): editorial parameters; an assisted answer carries no weight	`amp/src/learning/frontier-policy.ts` (lines 175–214)
[S7]	Practice target: working level minus half a grade plus 0.3 times its uncertainty	`amp/src/learning/practice-policy.ts` (line 207)
[S8]	Batch selection: targeting, support work and variety; at most five problems a batch	`amp/src/server/mixed-practice.ts` (line 153)
[S9]	Documentation: the production estimator organizes practice and is not a validated placement	`amp/docs/adaptive-evidence.md` (line 6)

	What it shows	File at the cited commit
[S10]	Problem contract: worked explanation and feedback on a chosen wrong option	`amp/src/math/contracts.ts` (line 456)
[S11]	Practice screen: the result, part marks and worked explanation after every answer	`amp/src/web/PracticeWorkspace.tsx` (line 531)
[S12]	Answer checking: `gradeAnswer` sends each answer to the checker for its answer type	`amp/src/math/engine.ts` (line 762)
[S13]	Multi-part answers: each blank is marked on its own	`amp/src/math/multi-part-choice.ts` (lines 3–17)
[S14]	Step-by-step work: each step or proof line is marked on its own	`amp/src/math/structured-work.ts` (lines 44–63)
[S15]	Lessons: worked example, method and film for each problem type	`amp/src/content/problem-lessons.ts` (line 60)
[S16]	A problem type's lesson method: goal, plain-language intuition, glossary terms, steps, the error to watch	`amp/src/content/problem-lessons.ts` (lines 129–146)
[S17]	Lesson and problem content by topic	`amp/src/content`
[S18]	The shared glossary: term cards that words in hints and explanations open	`amp/src/glossary/index.ts` (line 15)
[S19]	Workbooks: printed work chosen from one learner's evidence	`amp/src/workbook/compose.ts` (line 15)
[S20]	Hint: one teaching step written from the problem's own givens, never the answer	`amp/src/math/contracts.ts` (line 442)
[S21]	Test: every problem type's instances carry a hint that names the first move and never states the answer	`amp/tests/instance-hints.test.ts`
[S22]	Skill status: review-ready after independent success on the latest attempts	`amp/src/server/progress.ts` (lines 17–42)
[S23]	Help events (lesson, hint with its step, worked example, video, print), their day and week rollups, and time on help surfaces	`amp/migrations/0111-help-events.sql`
[S24]	Prerequisite repair after repeated errors	`amp/src/learning/practice-policy.ts` (line 215)
[S25]	Spaced-retrieval intervals: six hours to thirty days	`amp/src/server/evidence.ts` (line 26)
[S26]	Film kit: a screenplay drawn live as an animated lesson, synchronized to narration	`amp/film/README.md` (line 3)
[S27]	Lesson films: one film per problem type, versioned and served beside its lesson	`amp/src/content/lesson-films.ts` (line 28)
[S28]	Film player: opening a film is recorded, and its time counts as video time	`amp/src/web/features/films/FilmScreen.tsx` (line 78)
[S29]	The film manifest: every published film and the problem types it serves	`amp/src/content/films/manifest.json`
[S30]	Video playback: a video played, and played to its end, once per opening, rolled into the day and the week	`amp/migrations/0145-learning-doses.sql` (lines 8–35)

	What it shows	File at the cited commit
[S31]	Family views: each child's progress record, workbooks and settings	`amp/src/web/features/family`
[S32]	Teacher views: daily needs list, class map, gradebook, help used, progress record	`amp/src/web/features/school/classes`
[S33]	Paper workbook results: typed answers graded at full weight, a right / not-yet mark at half weight	`amp/src/workbook/results.ts` (lines 26–27)
[S34]	`amp_adult_checkins`: an adult's view, assignment or workbook score, by role (parent, teacher, aide, school staff), per learner	`amp/migrations/0145-learning-doses.sql` (lines 41–49)
[S35]	Points: one XP for each correct answer without help, shown with the result	`amp/src/web/features/practice/AnswerFeedback.tsx` (line 64)
[S36]	Skills inventory: XP and a filling bar for each skill	`amp/src/web/features/inventory/InventoryScreen.tsx` (line 360)
[S37]	Short animations on a correct result and on a mastered skill	`amp/src/web/feedback-sheet.css` (line 176)
[S38]	Active time is recorded per learner, day, surface and skill; the surfaces: practice, lesson, feedback, map, other, hint, video, worked example	`amp/src/schools/evidence.ts` (lines 19–29)
[S39]	Skill-evidence summaries kept for every practised skill	`amp/migrations/0008-skill-evidence.sql`
[S40]	Growth: pace of the stored level on settled days, with its standard error	`amp/src/schools/growth.ts`
[S41]	Server-side crediting shared by every heartbeat: at most 90 seconds, and at most the time since the last one plus five seconds	`amp/src/schools/evidence.ts` (lines 74–92)
[S42]	Integration test: a child on a parent's account in no class gets exactly the rollups a school student gets	`amp/tests/every-learner-measured.integration.ts` (line 118)
[S43]	The production-calibration harness: simulated learners, including misspecified ones, run through the production estimator	`amp/scripts/frontier/production-calibration.ts`
[S44]	Problem generators and the per-format answer-checker modules	`amp/src/math`
[S45]	Problem flags: a learner or adult can report a problem	`amp/src/feedback/problem-flags.ts`