Pictures Are For Babies ESSA Tier 4 Report

Contents

Introduction

Passed in 2015, the Every Student Succeeds Act (ESSA) is a US law that governs K-12 public education policy. ESSA aims to provide all students with a fair, equitable, and high-quality education. One of the key components of ESSA is the classification of educational programs into tiers based on their effectiveness and evidence of impact.

Tier 4 programs are those that have not yet demonstrated evidence in well-designed and well-implemented studies but demonstrate a rationale for their potential effectiveness and are in the process of gathering evidence. This tier allows new and innovative programs to be recognized and be eligible for some federal funding while they continue to build their evidence base.

Pictures Are for Babies is a literacy software tutor designed to help students master every aspect of literacy, from alphabetic knowledge to reading and writing at the graduate and professional level. Given the extent of its scope, innovative technology, and uncompromising adherence to research-

based practices, it is important that educators and policymakers understand how *Pictures Are for Babies* is based on a rigorous logic model that meets all the requirements for ESSA Tier 4 compliance and how it plans to gather evidence of its effectiveness.

Executive Summary

Because of the length of this document, this section provides an executive summary that educators and administrators can use to quickly understand the program's logic model and its alignment with ESSA Tier 4 requirements.

- National and international assessments confirm that literacy instruction is failing at both ends of the distribution. Large numbers of students never reach basic proficiency, and the number of students demonstrating the highest levels of literacy is vanishingly small, with some assessments grouping the top two categories because so few students reach the top one. These assessments test only reading, so the number of students demonstrating the highest levels at both reading and writing is likely even lower.
- *Pictures Are For Babies* is a literacy software tutor that spans every aspect of literacy from alphabetic knowledge through graduate and professional reading and writing. It is powered by *Trane*, a deliberate practice engine that schedules, sequences, and tracks every exercise across five integrated tracks. Multi-tiered intervention protocols are built into the program to ensure that students receive the right level of support, and these protocols are designed to be implemented by any fluent reader willing to act as a tutor.
- The scientific consensus on how fluent reading develops is settled, but without a corresponding engineering effort, that consensus has failed to translate into classroom outcomes. An engineering of reading is an approach that designs solutions from the ground up to solve user problems rather than test hypotheses, to anticipate variation across learners and contexts rather than control for it, and to make errors the designer's responsibility rather than the teacher's. *Pictures Are For Babies* represents the most comprehensive attempt to put this approach into practice.
- The program's logic model is built on deliberate practice, the framework shown across domains from music to chess to medicine to produce expert performance. *Trane* automates the conditions for deliberate practice by implementing spaced repetition, interleaving, retrieval

practice, and mastery learning. These cognitive science components have empirical support and are integrated with the science of reading and the science of writing into a single coherent system. The program is not an attempt to introduce novel theories, but to integrate and operationalize decades of research into a system that can be implemented at scale.

- A session requires only a tutor, the software, a medium for dictation, and books of the student's choice. Lessons are either reading or dictation, applied across all the curriculum. In a reading lesson the tutor reveals hidden examples and checks whether the student reads them fluently. In a dictation lesson the tutor reads examples aloud while the student writes them down. If the student makes mistakes, the tutor performs an intervention to explicitly teach the material the student missed. Reading breaks let the tutor read aloud from a book of the student's choice, giving them a chance to rest while continuing to develop background knowledge and vocabulary. A new tutor can be trained to use the software in fifteen minutes and on the intervention protocols in a few hours.

- Drawing from Kilpatrick's categorization of reading interventions, the program follows the pattern of the most effective programs ever documented: explicit phonics instruction, phonemic awareness developed to proficiency, and extensive practice in connected text. It extends this pattern of uncompromisingly using the best research to include reading comprehension through building background knowledge and explicit writing instruction from sentences to multiple paragraphs through worked examples.

- The expected outcomes for graduates of the program are not mere promises but guaranteed by its design. Because completion requires demonstrated mastery at every level, a student who finishes the program must possess fluent and accurate reading and spelling of most words in the English language, the ability to comprehend undergraduate and professional-level texts, a strong mental model of how to write at every level from sentences to multi-paragraph texts, and enough writing practice to have transferred this model to their own writing.

- The program anticipates failure modes and implements structural safeguards that shift responsibility for errors from teachers to the engineered system. Tutors require no specialized training, software complexity is hidden behind a simple interface, human tutors handle speech and dialect variation, break-resistant design prevents gap accumulation, and eschewing gamification in favor of deliberate practice.

- The research consensus that 90 to 95 percent of students can learn to read proficiently with effective instruction provides a benchmark for what is possible. Estimating an effect size from that finding and the current performance of 12th graders, any program that closes the gap would be unprecedented in the history of literacy education. Students deserve such an intervention, and failing to engineer it is a failure to meet their needs and capacities.
- The author is seeking an independent research partner to conduct a multi-year longitudinal randomized controlled trial. Full access to the program will be provided at no cost to researchers, tutors, and participating schools.

Logic Model

Problem Statement

Despite overwhelming scientific consensus on the mechanisms that drive reading and writing acquisition, actual instruction often fails to align with its recommendations. Even in programs that align with research, much of its implementation is left to educators, who cannot reach fidelity even with extensive training. The state of literacy instruction is even more dire for struggling students, who require extensive individualized instruction. In the absence of what Elizabeth Tipton and Nicole Patton-Terry call an "engineering of reading", the "science of reading" is no more than a marketing badge that fails to translate to real outcomes even when the programs align with its recommendations.

National and international assessments of literacy confirm that literacy instruction is failing to meet the needs of students. The OECD Survey of Adult Skills (PIAAC) shows that 28% of adults have low literacy proficiency. The Program for International Student Assessment (PISA) shows that 26% of students in all member countries score below level 2 in reading, with the United States having 19% of students in this category. The National Assessment of Educational Progress (NAEP) shows 32% of 12th graders in the United States scoring below basic in reading.

These figures are often the focus of the discussion, but the results paint a more dire picture when focusing on the highest-performing students. Across all three assessments and all countries, the number of students demonstrating the highest levels of literacy is under 10%, with some assessments grouping the statistics for this group with the next highest level because the number is so small as to be meaningfully reported as its own category. These assessments test only reading performance,

so the number of students demonstrating the highest levels at both reading and writing is likely even lower.

Pictures Are for Babies is not designed for mere remediation of low-performing students or to meet some government standard. From its very inception and through every aspect of its design and implementation, it aims to engineer the conditions that make the highest levels of literacy the default for all students that complete it. In light of the continued challenges in achieving these outcomes, this goal might seem impossible. However, as Louisa C. Moats states in her seminal article "Teaching Reading Is Rocket Science", "reading failure can be prevented in all but a small percentage of children with serious learning disorders. It is possible to teach most students how to read if we start early and follow the significant body of research showing which practices are most effective".

The rocket science of literacy instruction is settled, but without the corresponding aerospace engineering, the spaceship will not reach its faraway destination. What Tipton and Patton-Terry describe as the "engineering of reading" is an approach that designs solutions from the ground up to solve user problems rather than test hypotheses, to anticipate variation across learners and contexts rather than control for it, to make errors the designer's responsibility rather than the teacher's, and to be refined through iteration rather than validated through a single study.

Unbeknownst to Tipton and Patton-Terry, the author, working independently and without institutional support, had already completed much of what they call for at the time of publication. That this system was built by a single person is not a coincidence. An effort this uncompromising could only have emerged outside the constraints of edtech product committees, venture capital, and market-driven consensus. The forces that prevent established players from integrating research at this depth are the same forces that prevent them from naming a literacy program *Pictures Are For Babies*, let alone commit to the pedagogical and engineering decisions such an ambitious goal requires.

Logic Model Summary

Pictures Are For Babies is built on a foundation of best-in-class pedagogy, drawing from a multidisciplinary body of research in cognitive science, expert performance, optimal learning, the science of reading, and the science of writing instruction. These disparate fields not only align, but reinforce each other, forming a coherent theoretical and practical framework for effective literacy instruction.

Unfortunately, while the research has been replicated and validated for decades, classroom instruction and most commercial programs are, at best, decades behind, and, at worst, actively hostile to it. Some programs correctly and thoroughly execute parts of the research, but none can claim to integrate all of it into a coherent whole. *Pictures Are For Babies* is the first program to do so. Before detailing how every aspect of the program is grounded in research, this section presents a high-level summary of the logic model that underpins it, both to provide a short summary for policymakers and educators, and to provide full context for the detailed logic model that follows.

The backbone of the logic model is the program's focus on deliberate practice. Deliberate practice is a well-established framework for explaining how individuals acquire expertise in any domain. This framework has been validated in a wide range of domains, from music to chess to sports to medicine. It states that expert performance across any domain is achieved through effortful, structured training with immediate feedback aimed at improving specific aspects of performance. Deliberate practice is not meant to be fun, inherently enjoyable, or easy. Deliberate practice succeeds because it relentlessly targets weaknesses at the edge of current ability and develops the learner's ability to self-monitor and self-correct, and fosters their intrinsic motivation to improve.

Pictures Are For Babies implements deliberate practice through *Trane*, an engine that creates the conditions for deliberate practice in any domain. It surfaces optimized and personalized exercises based on the learner's past performance, a curriculum designed to target individual skills, and the dependence relationships between these skills. *Trane* uses many well-established techniques from cognitive science, such as spaced repetition, interleaving, and mastery learning that will be discussed in detail in the next section. *Trane* was designed from scratch by the creator of *Pictures Are For Babies* and validated in other fields like music before being applied to literacy instruction. Deliberate practice is not a buzzword or a retrofit, but core to the design and implementation of *Pictures Are For Babies*.

Existing literacy research and recommendations are well aware of the importance of deliberate practice, but the existing solutions have so far been unable to implement it at scale beyond simple adaptive algorithms and self-paced learning models. Without a proper and strict implementation of deliberate practice, we can expect that the number of students that reach the highest levels of literacy will remain very low. No other model of expert performance has been shown to be as effective as deliberate practice, so expecting students to reach those levels is akin to expecting novice pianists to become concert pianists by publishing graded sheet music without years of extensive

structured practice and feedback.

This paradigm works much better when the instruction and feedback are individualized to the learner. *Trane* solves the problems of providing individualized instruction at scale. *Pictures Are For Babies* goes further by implementing a model of reading and writing instruction that is both scientifically grounded and delivered through simple protocols any fluent reader can follow. Through this combination, teachers, parents, and volunteers can provide individualized instruction to students without needing to be experts and without needing to know the prior knowledge of the student. Schools can thus implement both small group and one-on-one instruction without needing to hire additional staff or provide extensive training.

By treating deliberate practice as an engineering problem, *Pictures Are For Babies* is able to implement and enforce the science of reading and writing instruction in a way that is impossible for human instructors and printed materials to do. For example, by keeping track of the performance of every word lesson, the software can identify when a student should receive a higher tier intervention and alert the tutor. While a teacher might be able to identify that a student is struggling, *Pictures Are For Babies* can perform this analysis with precise performance data and at the lesson level on a word curriculum that spans over 17,500 words. In another example, the concept of Scarborough's reading rope explains how the many different skills that contribute to literacy acquisition are interdependent and interact. Many products pay lip service to this concept but fail to implement it in practice. In *Pictures Are For Babies*, the concept is implemented without any human intervention because *Trane* makes sure that the exercises shown to the student come from a variety of courses and lessons testing different skills. Many more of these examples will be discussed in the next section.

Beyond allowing for implementing and enforcing the science of reading and writing instruction, *Trane* also allows *Pictures Are For Babies* to implement a curriculum that is far more extensive than any other program and to thoroughly adhere to the science without fear that it will require large amounts of effort on the part of the tutor. The curriculum of *Pictures Are For Babies* covers every aspect of literacy, namely:

- A symbols track that teaches alphabetic knowledge and knowledge of basic punctuation and symbols.
- A words track that teaches word-level reading and writing. By having *Trane* take care of spaced review and interleaving, the words track can cover over 17,500 words with a simple

intervention protocol that does not require the tutor to have any knowledge of linguistic concepts, detailed scripts, or lesson plans. The only requirements for a tutor are to be a fluent reader, the phonetic skills all fluent readers possess, and the willingness to learn and follow this simple protocol.

- A sentences track that teaches sentence-level reading and spelling and implicitly introduces syntax and grammar. Each sentence course follows a corresponding word course and is designed to reinforce the words learned in that course.
- A reading comprehension track that teaches reading comprehension through a combination of reading a large variety of texts of many types, genres, and topics and developing the background knowledge necessary to understand them. The texts in this track are carefully selected and annotated with example questions for the tutor and footnotes that provide additional context and background knowledge.
- A writing track that teaches writing by having students study and analyze worked examples of explicit writing instruction exercises at the sentence, paragraph, and multi-paragraph level. Each writing course follows a corresponding reading comprehension course and its material is derived from the texts in that course to further reinforce it. Writing production is outside the scope of the deliberate practice framework for reasons explained in the next section, but tutors can use the writing exercises to provide writing production practice and feedback to students.

In terms of adherence to the science of reading and writing instruction, *Pictures Are For Babies* follows all the standard recommendations for literacy instruction. These choices are thoroughly explained in the next section, but a few examples include:

- Systematic and explicit instruction of alphabetic knowledge, phonemic awareness, and phonics.
- Teaching spelling along with reading to reinforce each other.
- Reading comprehension instruction based on the simple view of reading by teaching decoding and language comprehension skills.
- Explicit writing instruction that teaches students how to write at the sentence, paragraph, and multi-paragraph level.

In order to create a system that engineers the conditions for expert literacy performance in most students, it is necessary to consider the needs of every student. For the lowest-performing, the

scientific base has shown that reading difficulties mostly stem from deficits in phonological processing, and that with early and adequate intervention, over 90% of students can achieve grade-level reading proficiency. By focusing its multi-tiered word course intervention protocol on developing the required phonological processing skills, *Pictures Are For Babies* helps the students with deficits in this area. For students with deficits in other areas, professionals can use *Pictures Are For Babies* to provide individualized instruction and deliberate practice while they focus on delivering the specific interventions that these students require.

For average students, *Pictures Are For Babies* ensures that any gap in skills is addressed before it becomes a larger academic problem. High-performing students, who are often neglected by existing group instruction, can use the scalable individualized instruction to continue to develop their skills without regard to grade levels and without worrying that the cost of this level of instruction will siphon resources away from other students. In this way, *Pictures Are For Babies* can serve as a universal literacy intervention that handles the needs of all students, from those with learning and phonological deficits to the highest-performing group.

Inputs and Activities

Having described the problem and the logic model, this section details the inputs and activities involved in *Pictures Are For Babies* by describing how a study session is conducted. Apart from the student, the tutor is the most important input to the program. Tutors are tasked with operating the software (thus removing the cognitive load of operating it from the student), checking off the examples that the student mastered in every lesson, conducting interventions when the student misses answers, and conducting reading breaks when the student is frustrated or bored. Apart from the tutor, student, and the software, the only other inputs are a medium for the student to take dictation and books of the student's choice for reading breaks.

The main activity is conducting lessons, which are of two types: reading and dictation. In a reading lesson, the screen presents the tutor with multiple hidden examples of material. The tutor reveals each example, asks the student to read it aloud, and checks off whether the student read it fluently. The criteria for fluency differ slightly depending on the material, but in general, the student is to read the material accurately and to start reading fluently within one second of seeing it. Longer material adds additional criteria, such as reading with appropriate prosody and phrasing and comprehension checks. In a dictation lesson, the tutor is again presented with hidden examples.

The tutor makes sure the student cannot see the examples, reveals each example, and reads it aloud to the student. The student is to write down what they hear, and the tutor checks off whether the student wrote it down correctly and fluently. Dictation lessons are not handwriting lessons, so students can use any medium to write down the examples, and younger students can vocally spell the answers.

Before moving to the next lesson, the tutor performs an intervention if the student missed any examples. The specifics of the intervention depend on the material and type of lesson, but they generally focus on explicitly teaching the student the material they missed. The more complicated word intervention protocol has multiple tiers that are triggered on repeated difficulty and involve procedures that train the student's phonological processing skills. Warnings for these higher-tier interventions are surfaced in the user interface so that tutors do not have to wonder when to perform them. Tutors may also initiate them proactively if they already know the student needs the extra practice.

These two types of lessons are applied to the five tracks of the program: symbols, words, sentences, reading comprehension, and writing, with dictation lessons only being used in the first three tracks. In this way, the program implements a complete literacy curriculum that covers every aspect of literacy without ever resorting to superfluous activities that do not match their real-world counterparts.

The second activity in the program is the reading break. Given the requirements of deliberate practice and the lack of pictures and gamification, students can understandably become frustrated or bored with the program. Instead of offering them an unrelated activity or game, the tutor and student can decide together to take a break. During this break, the tutor reads aloud a book of the student's choice, ideally letting the student follow their finger as they track it on the page. After the break, the tutor and student can decide to continue with the lesson or end the session. Because of *Trane*, the session continues right where they left off.

Detailed Logic Model

This section details the research and evidence that underpins the design and implementation of *Pictures Are For Babies*. It is organized into multiple sections for each of the fields that inform the program's design, and each section shows a principle from the research followed by an explanation of how *Pictures Are For Babies* implements it. For simplicity, *Pictures Are For Babies* will be

referenced as *PAFB* in the rest of this section. *Trane* is the name of the practice engine that powers *PAFB*.

Cognitive Science

- Learning and immediate performance are not the same. Immediate performance in a task is not a guarantee of long-term retention or transfer of that skill.
  - *PAFB* asks students to repeat the same task multiple times at optimal intervals. Mastered exercises are shown occasionally to make sure foundational skills remain solid.
- Working memory is the type of memory used for immediate performance. This type of memory is very limited in capacity and duration. Any type of instruction that requires working memory incurs a cognitive load that must be effectively managed.
  - *PAFB* follows this principle in many ways. For example, it scaffolds exercises to ensure only a small amount of new information is introduced at a time, uses a simple UI to minimize distractions, and does not teach complicated phonic rules by name.
- Just like working memory, attention is limited and must be guided.
  - *PAFB* uses a simple UI to minimize distractions. The sessions include explicit instructions to take breaks to replenish attention. Instructions guide students to focus on the correct aspects of literacy and avoid activities not backed by research.
- In contrast, long-term memory has a vast capacity and retains information for long periods of time. Complicated tasks are performed by moving building blocks to long-term memory. Very complex skills like literacy involve repeated instances of this process.
  - *PAFB* starts at the very beginning with learning to read and write single letters and very basic words. Once they are in long-term memory, sentences and more complex words are introduced. Then real texts and writing tasks are introduced in increasing levels of complexity.
- Subconscious processes do not require active effort and are performed much faster than tasks that require it. Moving information to long-term memory enables the skills to become automatic and free working memory and attention to work on more advanced skills.
  - *Trane* supports this process by providing spaced review and gating of progress until mastery of easier skills is proven. Together with the scaffolded lessons, it supports the process of turning new skills into automatic habits.

Expert Performance

- Deliberate practice is the essence of expert performance. It is effortful, structured training with immediate feedback aimed at improving specific aspects of performance. It is not meant to be fun, inherently enjoyable, or easy.
  - *PAFB* uses *Trane* to primarily surface exercises at the edge of the student's current abilities. The lessons break literacy into small skills and students receive immediate feedback on their performance.
- Experts continually seek to improve their performance instead of falling into automatic patterns.
  - *Trane* is always gently pushing students past their current edge of ability. Over time, students build on their current skills until they master more and more difficult ones.
- Experts organize skills into large chunks to process complicated information faster.
  - *Trane* allows skills to be split into smaller skills and to explicitly encode the relationships among them. As students progress through this connected graph, they are given tasks that require that they build these mental structures to demonstrate mastery.
- Expert performers engage in self-regulated learning, setting specific goals, monitoring progress, and adjusting strategies when necessary.
  - *Trane* tracks learner performance and adapts exercise difficulty accordingly. Tutors can guide students toward deliberate strategy use, especially when progress stalls.
- Experts possess strong metacognitive skills, allowing them to reflect on their performance and make effective adjustments.
  - *PAFB* trains students to identify confusion and rely on tutors to clarify misunderstandings, building foundational metacognitive habits early.
- Motivation plays a critical role in sustaining the effort required for long-term improvement.
  - *Trane* and the *PAFB* tutor model aim to foster intrinsic motivation by tying progress directly to increased competence and ensuring a clear sense of purpose in all tasks.
  - The manual allows the tutor to make use of reading breaks in which they read interesting texts to the student when they detect frustration or boredom. These reading breaks allow students to relax while reinforcing their love of literacy and developing their background knowledge and vocabulary.
- Experts develop flexible, adaptable skills that allow them to transfer knowledge across situations while still relying on deep domain-specific knowledge.

  - *Trane*'s graph structure allows students to master discrete skills and apply them in increasingly varied contexts, from decoding novel words to understanding complex texts and expressing themselves in writing.
- Early advantages and perceived talent are not predictive of long-term outcomes. Students who appear "gifted" early on often plateau without the right kind of input, while those who build strong foundations through sustained, structured, and deliberate practice eventually surpass them.
  - *PAFB* produces this kind of quality input for all students, regardless of initial ability. It does not cater to "gifted" early learners or the fastest students. Instead, it supports all students, providing the same rigor, structure, and opportunities for growth. The only difference is individual pace, adjusted based on mastery, not superficial measurements of early ability. Excellence is developed over years, not discovered or accelerated in a race for early achievement.

Optimal Learning

- Mastery learning ensures that students do not progress until they have demonstrated mastery of the current material. This prevents gaps from accumulating and supports long-term growth.
  - *Trane* automatically withholds access to more difficult material until the student has demonstrated consistent success with prerequisite skills. Mastery is based on performance, not on completion.
- Spaced repetition improves retention by revisiting material in optimal intervals, just before it is forgotten.
  - *Trane* uses time-aware review scheduling to resurface lessons at optimal intervals based on each student's history, ensuring long-term retention without overloading them with review.
- Retrieval practice is the act of recalling information from memory, and it has been shown to be more effective than reviewing material or taking notes. The act of trying to recall information strengthens memory and highlights gaps in knowledge.
  - All *PAFB* exercises are a form of retrieval practice. Students are first asked to recall how individual words are read and spelled. Higher level tasks require recall of individual words, background knowledge, word meaning, and text structures.
- Interleaving different types of tasks strengthens retention and transfer by forcing students to

repeatedly retrieve and distinguish between skills.

  - *PAFB* implements interleaving through randomized review selection and frequent alternation between decoding, spelling, sentence writing, and comprehension tasks.

- Mixing tasks of varying difficulty improves engagement and learning efficiency. Too much difficulty leads to failure and too much ease leads to stagnation.
  - *Trane* surfaces a blend of easier and harder tasks based on the student's performance scores, maintaining an optimal difficulty range for learning.
- Implicit repetition strengthens learning by increasing practice on related tasks when a student struggles, and decreasing practice when a student does well.
  - *Trane* implements a reward system. When a student performs well, simpler and related skills are given a positive reward so that they are practiced less often. When a student performs badly, related, more complicated skills are given a negative reward so that they are practiced more often.
- Flow is a state of deep concentration and enjoyment that occurs when a task's difficulty is well-matched to the learner's current skill level. It is characterized by sustained focus, loss of self-awareness, and high intrinsic motivation.
  - *PAFB* and *Trane* use all the strategies above to best match the exercises to the student's current abilities. Flow is the result of implementing optimal learning strategies, not of turning learning into gamified experiences or hacking motivation.
- Accelerating the timeline for mastering a complex skill is possible, but only under specific conditions. Research shows that true acceleration does not come from shortcuts, gamification, or early performance boosts. Rather, it comes from compressing high-quality, representative experience into tighter timelines using structured feedback, gradual complexity, and expert-like modeling.
  - Similarly to how *PAFB* achieves flow, acceleration is not a goal in itself, but the result of implementing optimal learning strategies, combining them with the best reading pedagogy, and a deep and wide curriculum that progressively exposes learners to all aspects of literacy.

The Science of Reading: Word and Sentence Reading and Writing

- Phonological awareness is the ability to recognize and manipulate the sounds of spoken language. Phonemic awareness is a more specific skill involving the ability to isolate and work

with individual phonemes, the smallest units of sound.

  - *PAFB* builds these skills from the beginning. Students learn to hear and manipulate sounds in words before mapping them to letters.

- Most reading difficulties are caused by weaknesses in phonological processing, not vision or motivation. Students who cannot hear, segment, or manipulate sounds will struggle to decode and map words into memory.

  - *PAFB* targets these core deficits directly. Every word and sentence lesson includes tasks that strengthen the phonological system and link it to written language. No guessing strategies or visual cueing systems are used.

- Orthographic mapping is the central process that allows readers to store words in memory for instant recognition. It links the pronunciation, spelling, and meaning of words, and depends on phonemic awareness and letter–sound knowledge. Without successful orthographic mapping, students will not become fluent readers, no matter how much they are exposed to print.

  - *PAFB*'s lessons on words and sentences are built on this concept. Tutors are instructed to mark words and sentences as mastered only when reading and writing are effortless and automatic.
  - *Trane* ensures that progress is gated behind accurate mapping of the words, not on mere decoding or guessing.

- Orthographic mapping should extend beyond simple words to include irregular spellings, multisyllabic words, and morphologically complex forms. Fluent reading requires rapid retrieval of tens of thousands of words, including those not easily decoded by rule.

  - *PAFB* is designed to support deep orthographic mapping. The system introduces increasingly complex word forms only after foundational skills are in place and goes all the way to ensure students master words needed to proficiently read undergraduate and graduate material.
  - The word curriculum of *PAFB* covers over 17,500 unique words. Through repeated exposure to real texts in more advanced courses, their vocabulary continues to grow beyond the words explicitly taught.

- The self-teaching hypothesis posits that once a student develops efficient decoding skills, successive encounters with new words allow them to apply their existing knowledge to map them in long-term memory with fewer trials. Through this process, fluent readers develop a sight vocabulary that contains tens of thousands of words.
  - *PAFB* supports this process by providing ample opportunities for students to encounter words in varied contexts, promoting deeper understanding and retention, and enabling generalization of spelling patterns. As students progress through the word curriculum, they steadily use their previous knowledge to expand their sight vocabulary.
  - Unlike the raw application of the hypothesis, *PAFB* does not leave this to chance. The curriculum and the practice engine ensure students efficiently develop a large sight vocabulary.
- With proper instruction that targets phonemic skills and supports orthographic mapping, the vast majority of children can learn to read proficiently. Research estimates that 90 to 95 percent of children can reach grade-level reading ability with effective interventions.
  - *PAFB* is built on the methods shown to produce these outcomes. By addressing phonemic awareness, decoding, spelling, and comprehension in the correct order, it ensures that reading failure is not treated as inevitable, but as preventable.
- A hallmark of the most successful interventions is teaching phonemic awareness to proficiency in students that exhibit persistent reading difficulties, including tasks like phoneme deletion, substitution, and reversal. These skills are essential for supporting orthographic mapping and are often missing in students with persistent reading difficulties.
  - *PAFB* gives explicit instructions on how and when to address deficits in phonemic awareness with these advanced tasks.
- Research supports a tiered model of instruction, where all students receive core instruction and only those who struggle are given additional support.
  - *PAFB* follows this model closely. Every student receives explicit, structured lessons, but when a student struggles with a task, additional instruction and targeted phonemic work are introduced. This ensures that time is spent efficiently and support is provided only where needed.

  - *PAFB* uses the student's performance data to determine when to suggest these higher tier interventions and shows a graphical warning.

- Phonics is essential for teaching how letters represent sounds, but it is not enough to produce fluent readers. Letter-sound knowledge must lead to permanent connections between sounds, spellings, and meanings.
  - *PAFB* uses phonics to organize instruction and support orthographic mapping, but does not teach rules explicitly. Concepts like vowel types or syllable labels are avoided to reduce cognitive load. Every word, regular or irregular, is taught by mapping its phonemes to its spelling, and no word is ever treated as an unbreakable whole to be memorized.

- Phonemic awareness instruction works better when it is integrated with phonics instruction, not as standalone units that precede phonics instruction by months or years. On the other hand, some like Kilpatrick argue that introducing letters too early can interfere with phonemic awareness development.

- *PAFB* solves the tension between these two positions by teaching phonemic awareness first followed by teaching the mapping of letters to sounds within the same lesson. That way, both skills are developed in parallel while allowing the tutors to notice deficits in phonemic awareness.

- The higher tiers of the intervention model are designed to address these deficits, and tutors are instructed to only use sounds and tokens in deletion, substitution, and reversal tasks. Without the use of letters in these tasks, students are forced to rely entirely on their phonemic awareness skills.

- Reading and spelling are two sides of the same process, each reinforcing the other. Spelling is more difficult because it demands exact recall of letter sequences.
  - *PAFB* treats spelling as a core component of literacy. Each word and sentence reading lesson is followed by a lesson where students must correctly write the words or sentences from the tutor's dictation.

- Word-level spelling develops along a predictable path. Early attempts are phonetic but imprecise. Over time, students learn to spell words more accurately by internalizing common letter patterns and morphological endings.

  - *PAFB* follows this trajectory deliberately. Students begin with simple one-syllable words and progress toward irregular spellings and affixed forms once the foundations are secure.

- Sentence-level reading and writing introduce new challenges. Students must coordinate multiple word forms, apply grammar, and maintain fluency.
  - *PAFB* begins sentence work only after students can reliably decode and spell the words involved. Sentences are used to reinforce word reading in context and to teach syntax and grammar through reading and taking dictation.

The Science of Reading: Comprehension

- The Simple View of Reading defines reading comprehension as the product of two components: decoding and linguistic comprehension. If either is weak, comprehension will suffer.
  - *PAFB* ensures decoding is fully mastered before comprehension is emphasized. Real texts are only introduced after the student can reliably read and spell all the words involved.
  - *PAFB* tackles linguistic comprehension by annotating texts with footnotes and briefings, with tutors available to cover any gap.
- Most comprehension problems in struggling readers are caused by poor word reading, not poor reasoning.
  - *PAFB* follows this principle by focusing early instruction on decoding, spelling, and syntax. Comprehension tasks are introduced only after students can fluently access the text.
- Vocabulary and background knowledge are critical predictors of comprehension. Readers must know the meaning of the words and the context behind the ideas they are reading.
  - *PAFB* includes a wide range of text types to develop both vocabulary and domain knowledge. Students are exposed to literature, nonfiction, poetry, essays and more, curated and leveled to support gradual knowledge growth.
- Skilled readers construct a mental model of the text by integrating ideas across sentences. This process depends on decoding, vocabulary, working memory, and attention.
  - *PAFB* supports this process through carefully leveled texts. As students progress, the texts become longer and more conceptually demanding, giving them practice with sustained reading and inference-making.

- Good readers monitor their comprehension as they read and notice when something does not make sense.
  - *PAFB* trains students to say "I don't know" when confused. Tutors are instructed to stop, clarify, and re-read as needed, helping students develop this metacognitive skill.
- Reciprocal teaching is a well-researched method for improving comprehension, especially in struggling readers. It involves the tutor modeling what skilled readers do when they encounter confusion by pretending not to understand and asking the student to explain or clarify.
  - *PAFB* uses this approach by having tutors ask questions aloud during reading, even when they know the answer. This strategy is used only when the tutor suspects the student may be confused, and is faded over time as the student becomes more proficient.

The Science of Writing

- Reading and writing reinforce each other. Reading builds the vocabulary, syntax, and background knowledge that writing draws on, while writing strengthens understanding of how texts are structured, and exercises word and sentence reading and spelling skills.
  - *PAFB* treats reading and writing as interdependent skills. Each level of text reading courses is followed by a writing course that introduces students to increasingly complex composition tasks.
- Writing is not a natural skill. It must be taught explicitly, in a structured progression, beginning with sentence construction and extending toward paragraphs and full texts.
  - *PAFB* and *Trane* are built on small and repeatable exercises that are easily verifiable by a tutor. Writing production tasks do not fit this mold. They cannot be repeated many times without boring students, take much longer than reading and spelling tasks, and have open-ended answers that would require more training and effort from the tutor.
  - This problem is solved by turning writing production tasks into reading tasks. Instead of asking students to solve writing tasks, *PAFB*'s writing track shows students how expert writers tackle the problems by having them read the problem and the solution.
  - The strategy makes the exercises repeatable, shorter, and removes their open-ended nature. Writing production tasks must be undertaken outside the program. *PAFB* supports production by exposing students to the same type of tasks and letting tutors use the writing questions as a starting point for their own production tasks.
- Worked examples are a powerful tool for teaching in a variety of subjects, as they provide

students with a clear model of how to approach problems that they can then generalize and emulate.

  - Including writing production tasks in *PAFB* would have been ideal, but the decision to turn these tasks into reading of worked examples is not without precedent. The evidence shows that worked examples can be more effective than problem-solving, especially for novice learners.

- Writing instruction is most effective when directly tied to reading content. Disconnected prompts do not build transferable skills. The content of the curriculum drives the rigor of the writing activity.
  - Whenever possible, *PAFB* links each writing task to the texts from the previous reading comprehension course. The model answers reuse the vocabulary and ideas from those texts, helping students internalize structure and content together.
  - As the texts become more complex, so do the writing tasks associated with them.
- Students must learn to write complete sentences before they can write meaningful paragraphs or essays.
  - *PAFB* begins with structured sentence tasks that train students to express relationships clearly and precisely. Students are shown how to use because–but–so, appositives, and sentence combining to add depth and control to their writing.
- Writing tasks at the paragraph and text level must be addressed after sentence-level writing. The two most important phases of the writing process are planning and revising.
  - *PAFB* teaches planning by showing examples of structured outlines. These outlines make the internal structure of writing visible and teach students how to organize ideas before writing begins.
  - Revising exercises show what well-revised writing looks like. Tutors are encouraged to ask students how the response could be improved, giving them an opportunity to reflect on clarity, logic, and structure without having to generate a full rewrite.
- Grammar and mechanics should be taught in the context of actual writing, not through isolated drills.
  - *PAFB* exposes students to correct usage by having them read and study well-constructed sentences. Each task models how grammar supports clarity and meaning in real communication.

Expected Outcomes

Another advantage of using *Trane* in *Pictures Are For Babies* is that its mastery learning model allows us to make stronger assertions about the expected outcomes of students that complete the program than programs that lack its features. After all, a student can only be shown exercises from the very last courses in the curriculum if they have shown mastery of all previous courses. Before proceeding, it is important to mention that while the symbol, word, and sentence courses are complete, the reading comprehension and writing courses are still being developed and only the courses spanning early grades have been released. However, the missing courses (spanning from those early grades to graduate and professional material) are not expected to change the design and software of the program in any way. At the current pace of development, the full curriculum is expected to be completed and released in one to two years from the time of this writing (mid 2026). Therefore, all the assertions in this section are made about a complete version of the program. With this out of the way, we can say that students that complete the program must possess:

- Knowledge of every letter and common symbol in the English language, including their names, sounds, and common usages.
- Fluent and accurate reading of most words in the English language, including irregular spellings, foreign loan words, proper nouns, and morphologically complex forms. A sight vocabulary of this size is present in the best adult readers and is necessary for all other literacy skills.
- Fluent and accurate spelling of that same sight vocabulary.
- Fluent and accurate reading and spelling of those words in connected text.
- Ability to comprehend a wide variety of texts, including literature, nonfiction, poetry, essays, plays, scientific articles, journalistic writing, and more. This level of comprehension is not enabled by cookie cutter comprehension strategies, but by excellent word-level reading and the background knowledge required to understand their content and conventions, in line with the Simple View of Reading.
- Ability to comprehend such texts at the undergraduate, graduate, and professional level. These texts require that the reader integrate information across multiple dense and often conflicting sources, evaluate the strength of evidence and the credibility of arguments, infer meaning from indirect or embedded information, recognize the author's purpose, tone, and stylistic choices, and apply prior knowledge to interpret novel or abstract ideas. These skills

are what national and international assessments define as needed to be included in the top reading proficiency levels.

- A strong mental model of how to write at the sentence, paragraph, and multi-paragraph level. This model contains the knowledge of how to use conjunctions and transitions to chain sentences together, how to perform revision of single sentences, and how to plan and revise single paragraph and multi-paragraph texts, and more.
- Those students that complete writing production tasks outside the software with the aid of their tutor and the writing exercises will have ensured the mental model has been transferred to actual writing production.

It is still an open question to what extent students that do not complete writing production tasks outside the software are able to transfer the mental model developed through the writing courses to their writing production. Research on the reading-writing relationship shows that reading comprehension and writing quality are moderately correlated (r = .44), with skilled readers consistently outperforming weaker readers on writing tasks. However, the reading measured in these studies is normal independent reading, not reading of extensively worked examples showing the mechanics of the writing process delivered through a deliberate practice engine. Therefore, it is possible that the correlation between reading and writing in graduates of *Pictures Are For Babies* is much higher than the correlation found in the general population.

Estimated Effect Size Estimating the expected effect size of *Pictures Are For Babies* is impossible without conducting a large-scale randomized controlled trial. However, given that multiple studies suggest that 90-95% of students can learn to read proficiently at grade level with effective instruction, we can use this figure to estimate the expected effect size of any intervention that achieves near-universal levels of proficient reading.

The NAEP Proficient level is defined as solid academic performance and competency over challenging grade-appropriate subject matter. If we assume that proficient reading corresponds to this level, we can compute the effect size required to shift from the current 35% of 12th graders at or above Proficient to 90-95%. The NAEP Proficient cutoff sits at roughly 0.385 standard deviations above the general mean. The resulting effect size d for a given proportion p of students at or above Proficient can be computed using the following formula:

$d = 0.385 + \Phi{-}1(p)$

where $\Phi-1$ is the inverse cumulative distribution function of the standard normal distribution. These calculations assume that reading scores are approximately normally distributed and that the treatment and general-population distributions have equal variance. In reality, an intervention that moves a very large proportion of students above the proficiency threshold may produce a treatment distribution with lower variance, which would decrease the implied effect size. These calculations should therefore be interpreted as estimations.

A value of 90% of students at or above Proficient corresponds to an effect size of 1.67 and a value of 95% corresponds to an effect size of 2.03, for an average of 1.85. Any literacy intervention that produces an effect size of this magnitude would be unprecedented, but the continued failures of the past cannot be used as an excuse to stop trying. Failing to engineer such an intervention is a failure to meet the needs of the vast majority of students. This estimated effect size is also close to Bloom's original 2.0 effect size of mastery learning and one-to-one tutoring, the very intervention that *Trane* is designed to automate and scale.

Having shown that a literacy intervention with an effect size close to 2.0 is theoretically possible, we can examine how close *Pictures Are For Babies* comes to this idealized intervention. While *Pictures Are For Babies* claims to be an innovative and effective literacy intervention, none of its innovation lies in inventing or developing new methods of literacy instruction. Instead, it is a carefully designed and engineered synthesis of the best practices across the fields of cognitive science, expert performance, optimal learning, and the science of reading and writing. Therefore, while it is impossible to provide a direct estimate, we can examine the known effect sizes of the individual components of the program. These effect sizes are shown in the list below and sources are provided in the table and the Annotated Bibliography section to avoid cluttering the text.

- The most famous and one of the most controversial of these studies was published by Benjamin Bloom on the effects of mastery learning and one-to-one tutoring. Bloom found that students that received both interventions scored two standard deviations higher than students that received traditional instruction. Bloom's study proposes the famous problem that bears his name: finding a scalable solution that approximates the effects of one-to-one tutoring and mastery learning.
  - Attempts to replicate Bloom's results have never produced the same effect size, with recent meta-analyses of tutoring programs showing pooled effect sizes of 0.42 across all studies and 0.16 to 0.22 for large-scale programs using standardized tests. However,

the enormous labor costs of properly implementing one-to-one tutoring and mastery learning make it difficult to know whether the lower effect sizes reflect the ceiling of these interventions or the inevitable decay of implementation fidelity at scale.

- Mastery learning alone was found to have an effect size of 1.0 by Bloom. A meta-analysis found the overall effect size to be 0.59.
- Retrieval practice has an overall effect size of 0.50, with feedback increasing the effect size to 0.73.
- Interleaving shows an overall effect size of 0.42, with larger effects for more complex material.
- A meta-analysis of spaced repetition practice found an overall effect size of 0.54. Another one found an overall effect size of 0.74 when combined with retrieval practice.
- David Kilpatrick has categorized reading interventions into three tiers of increasing effectiveness. The most effective interventions are those that offer explicit instruction in phonics, develop phonemic awareness to proficiency, and present opportunities for reading in connected text. He identifies the second component as the secret ingredient that separates these interventions from less effective ones. Kilpatrick reports that these interventions result in a gain of 12-25 standard score points in nationally normed reading tests, with the majority falling in the 14-17 range, which is roughly an effect size of 1.0.
- A study evaluating Core Knowledge Language Arts (CKLA), a program that builds reading comprehension through a knowledge-rich curriculum, found an effect size of 0.47 on reading across grades 3-6.
- A meta-analysis of explicit writing instruction found that teaching explicit writing strategies produced an effect size of 1.02. A doctoral dissertation evaluating the sentence strategies outlined in *The Writing Revolution*, the specific writing strategies methodology used in *Pictures Are For Babies*, found an effect size of 0.84 in 9th graders. This dissertation has not been peer-reviewed, but it shows effect sizes within the range of the meta-analysis.
- A meta-analysis of worked examples versus problem-solving found an overall effect size of 0.57. Worked examples are more effective for novice learners because they reduce extraneous cognitive load.

Below is a table summarizing the expected effect sizes for interventions and instructional compo-

nents incorporated into of *Pictures Are For Babies*:

Component	Effect Size	Source
One-to-one tutoring + mastery learning	0.42 (at scale) to 2.0 (Bloom, narrow)	Bloom (1984); Kraft et al. (2026)
Retrieval practice with feedback	0.73	Rowland (2014)
Interleaving	0.42	Brunmair & Richter (2019)
Spaced repetition with retrieval practice	0.74	Latimier et al. (2020)
Advanced PA + phonics + connected text	~1.0	Kilpatrick (2015)
Knowledge-building comprehension	0.47	Grissmer et al. (2023)
Explicit writing strategy instruction	1.02	Graham et al. (2012)
Worked examples (vs. problem-solving)	0.57	Crissman (2006)

These components are not independent, and separate components can have synergistic and antagonistic effects. Therefore, an estimate of the actual effect size of an intervention that combines them cannot be derived by a mere sum. Otherwise, the effect size would be over 4.0 even with the most conservative estimates for mastery learning and one-on-one tutoring, which is clearly a ridiculous number. However, several synergistic effects are already known. Examples of synergistic effects include the integration of phonics and phonemic awareness, reading and spelling, and comprehension and writing. Given the breadth of research synthesized by *Pictures Are For Babies*, the integrated system may exhibit synergistic effects beyond those documented for individual combinations of instructional components.

Despite the impossibility of providing a precise estimate of the effect size of *Pictures Are For Babies*, listing the expected effect sizes of its components provides a strong indication that absent an unfaithful implementation of the research, the program should be able to produce a significant effect size. Whether the synergistic effects of combining these components will produce an effect size close to that of the idealized intervention is an open question.

A complementary way to estimate the expected effect size uses the expected outcomes described in the previous section and compares them with the highest levels of literacy measured in the National Assessment of Educational Progress (NAEP). In this assessment, students are classified into four achievement levels: Below Basic, Basic, Proficient, and Advanced, with only 5% of 12th graders reaching the Advanced level in 2024. Given that *Pictures Are For Babies* guarantees that completion and mastery are equivalent, that its curriculum consists of tasks representative of those required for real-world performance, and that the scope of the curriculum goes beyond the 12th grade, the program's design predicts that a significant proportion of a randomly selected group of students given access to the program and working through it in small-groups and one-to-one tutoring settings at school and at home will score at the Advanced level at the end of their K-12 education.

We can estimate the effect size implied by various proportions of *Pictures Are For Babies* graduates reaching that threshold using similar math to that used to estimate the effect size of the idealized intervention. The cutoff for NAEP Advanced corresponds approximately to the 95th percentile of the overall distribution, which sits at roughly 1.645 standard deviations above the general mean. The resulting effect size d for a given proportion p of PAFB graduates at or above Advanced can be calculated using the formula:

$d = 1.645 + \Phi{-}1(p)$

The table below shows the implied effect size for various proportions of *Pictures Are For Babies* graduates reaching the NAEP Advanced level, compared to the 5% observed in the general population.

% of PAFB graduates at NAEP Advanced	Implied effect size (d)
20	0.80
30	1.12
40	1.39
50	1.65
60	1.90
64	2.00
70	2.17
80	2.49

% of PAFB graduates at NAEP Advanced	Implied effect size (d)
90	2.93
95	3.29
99	3.97

Even conservative outcomes imply effects close to the largest components in the logic model. At just 20%, the number of graduates reaching NAEP Advanced would be four times the national average. If half of *Pictures Are For Babies* graduates reach NAEP Advanced, the implied effect size is 1.65, which exceeds every individual component in the summary table except Bloom's original finding. If approximately two-thirds of graduates reach Advanced, the implied effect size matches Bloom's 2.0 effect size. Larger values produce effect sizes that are difficult to imagine in a real-world setting, but that cannot be dismissed as impossible given that the difficulty of the curriculum compared to the test introduces ceiling effects.

Whether *Pictures Are For Babies* can achieve these outcomes and remove variability through its use of a deliberate practice engine and one-to-one tutoring is the central empirical question to be addressed by a future randomized controlled trial.

Research and Implementation Gaps in Other Literacy Programs

The main objection to believing that any literacy intervention can have the effect sizes needed to end the literacy crisis is that existing offerings have failed to do so, even those based on empirical research. Competing literacy programs often fail to incorporate all of these findings comprehensively or correctly, if at all. Rather than naming specific competitors and their shortcomings, this section focuses on the main misalignments between the pedagogy outlined in the previous sections and mainstream offerings. Parents and educators armed with this knowledge should be able to properly evaluate literacy programs and recognize those that truly align with best practices.

Cognitive Science, Expert Performance, and Optimal Learning

- Most programs make no attempt to incorporate findings from cognitive science, expert performance, or optimal learning. Concepts like spaced repetition, retrieval practice, cognitive load, or deliberate practice are not mentioned, much less implemented.

- There is no mastery tracking, no structured progression based on prior performance, and no attempt to match task difficulty to student ability. Tasks are selected based on curriculum scripts, not learning science.
- Even advanced programs that claim adaptability and some level of mastery learning do not follow the principles of optimal learning to their full extent. They use generic review schedules, present disconnected exercises, and treat success as one-time completion rather than repeated fluency over long periods of time.
- These are not minor omissions. They are a complete failure to build instruction on best practices.

Science of Reading: Word and Sentence Reading and Writing

- Most programs teach phonics but ignore orthographic mapping, the actual process that produces fluent readers. They encourage memorization of entire words, isolated rule drills, or guessing via context clues with no structured process to link sounds, spellings, and meanings in memory.
- Programs treat decoding as the criterion for success rather than fluent and effortless reading and writing. By doing so, they can hide reading difficulties in the short term and help students develop coping mechanisms that mask their struggles. In the long term, students with reading difficulties will hit a wall once texts become too difficult, abstract, and the pictures go away. This is a well-known phenomenon known as the fourth-grade slump.
- Phonemic awareness is treated as a warm-up or skipped entirely. Few programs include tasks like phoneme deletion, substitution, or reversal. These are essential for struggling readers and a hallmark of the most effective interventions.
- Spelling is often treated as separate from reading or not addressed at all. As a result, students are unable to internalize spelling patterns or apply them automatically.
- Sentence and text reading and spelling are introduced too early or not at all. Without a foundation of proper orthographic mapping, students will spend too many cognitive resources in word decoding and struggle with higher level tasks.

Science of Reading: Comprehension

- Comprehension tasks are introduced before decoding is fluent. This creates false positives, where students guess from context or pictures rather than read and understand the text.

- Most programs rely on comprehension strategy instruction rather than actual content understanding. Activities like predicting or summarizing are taught in isolation, without attention to vocabulary, syntax, or background knowledge, and without any link to the materials students are reading.
- Vocabulary and background knowledge are not systematically developed. Texts are often leveled to reduce complexity, not to build knowledge. Students read simplified content that does not prepare them for real academic reading.
- There is no clear pathway from decoding to deep comprehension. Programs treat word reading and comprehension as separate skills instead of building them in a structured, dependent progression.

Science of Reading: Composition and Syntax

- Writing is introduced through open-ended prompts with no modeling or structure. Students are asked to write before they've seen what good writing looks like. Sentence construction is rarely taught explicitly.
- Programs skip the sentence level and jump straight to paragraphs or essays. This overloads students and leads to incoherent writing. There is no foundation of grammar, logic, or structure to build on.
- Writing is disconnected from reading. Tasks do not reuse vocabulary or content from texts. There is no reinforcement between reading and writing activities.
- Planning and revision are not taught. Students are rarely shown how to structure a paragraph before writing, and revision is treated as fixing errors rather than improving ideas.
- Grammar is either not taught or reduced to isolated worksheets. There is no integration of grammar into actual writing, and no opportunity to see how sentence mechanics support meaning.

Partial Alignment

- Some programs incorporate parts of the research correctly. Some focus on early phonemic awareness and decoding. Others emphasize orthographic mapping, structured writing, or background knowledge.
- However, these components are rarely integrated into a complete system. Decoding may be taught without spelling, writing without sentence-level structure, or comprehension without

attention to background knowledge.

- Even when individual strategies are research-aligned, they are often isolated, loosely sequenced, or disconnected from a mastery-based progression.
- Even when the curriculum is perfectly designed and aligned with research, it is still the responsibility of teachers to plan lessons without the tools to track group and individual performance to select optimal exercises in ways aligned with best practices.

Some of the programs that meet partial alignment are listed below, along with the pieces they properly implement. Unlike the rest of the programs, these have made a concerted effort to align with the principles outlined in the previous sections, so deserve to be mentioned by name.

- *Heggerty*: Phonemic awareness instruction to the advanced level.
- *Equipped for Reading Success*: Advanced phonemic awareness in support of orthographic mapping, designed to capture the effects of the highly successful clinical interventions (*LiPS*, *Phono-Graphix*, *Discover Reading*) without requiring specialized clinical training.
- *Core Knowledge*: Developing the background knowledge and vocabulary necessary for comprehension.
- *The Writing Revolution*: Evidence-based writing instruction.

Failure Modes and Safeguards

Any attempt to engineer a literacy intervention that can produce the effect sizes needed to end the literacy crisis must be aware of the many possible failure modes and have undertaken measures to mitigate them. As Tipton and Patton-Terry (2026) argue, an engineering of reading assumes that students and tutors will make mistakes, and that the responsibility for mitigating them falls on the engineer, not on the users. *Pictures Are For Babies* takes responsibility as the detailed logic model demonstrates. This section lists these failure modes and makes the safeguards explicit.

- Teaching reading is a complicated job that requires a high level of expertise.
  - It is true that literacy education requires trained professionals as it stands. However, the responsibilities of a *Pictures Are For Babies* tutor have been reduced by completely taking care of sequencing, planning, and scheduling. The tutor's role is to score the student's performance, conduct interventions, and provide the emotional support that no machine can provide.
  - The intervention protocols are free from any linguistic or pedagogical knowledge require-

ments. Tutors are able to perform the interventions without any specialized knowledge, using only their reading skills and the phonemic awareness all fluent readers possess.

  - Fluent reading remains the scoring criterion throughout the program so that the tutors do not have to make subjective judgments about student performance. Simple rules like not allowing more than one second of hesitation before a word is read reduce variance.
  - Scoring of reading comprehension tasks is simplified by providing example questions for the tutor and having the student explain the text to the tutor instead of teaching specific strategies.
  - By employing worked examples in the writing courses, the tutor remains capable of guiding students even if they are not proficient writers themselves. Adapting the worked examples and using them as a scoring rubric is much easier than having to generate their own examples or evaluate student writing without a rubric.

- Even with most of the job automated, human tutors can fail to learn it or implement it correctly.
  - Special care has been taken to ensure that the intervention protocols are simple, clear, and do not require any specialized knowledge. The word protocol, the most difficult of them, is not based on having the tutor teach linguistic concepts or rules, or follow a predetermined script. Instead, the tutor uses their own phonemic awareness to demonstrate how it is applied to the words in the exercise and how those phonemes match to the letters in the words.
- Small group and individual tutoring is too expensive to implement at scale.
  - By making scheduling and lesson planning automatic, and by making the intervention protocols accessible to proficient readers, tutors can be recruited from a much larger pool of candidates. Schools can use parents, older students, teachers from other subjects, or even volunteers from the community to serve as tutors. A tutor can sit with a new student and instantly continue their education from where the last session left off.
- Tutoring the graduate and professional levels will be beyond the abilities of some tutors.
  - It is true that the later material will require a tutor with a matching level of reading proficiency. However, tutors that only have a high school reading level or lower can still be helpful to students that are in the early stages of the program. This lets schools allocate their more proficient tutors to the students that have reached those later stages. Parents that are not proficient readers can find a volunteer or a paid tutor (who does

not need to be a reading professional) to help them tutor their children through the later stages of the program.

  - *Pictures Are For Babies* is designed to be a lifelong literacy intervention. The intervention can also be applied to the tutors themselves to ensure they reach at least a high-school reading level. Research has shown that proper reading instruction has no age limit.

- The deliberate practice engine does not correctly implement the principles of optimal learning.
  - *Trane* is released as open-source software, meaning the code can be independently inspected and verified.
  - There are hundreds of tests that verify the correctness of the implementation. Because of the randomness of certain steps in the algorithm, there are also tests that simulate a student answering questions and verify the properties of the output. For example, a test verifies that a student that answers all questions correctly will see every exercise, while another verifies that a student that answers all questions in a course incorrectly will never move past it.
  - Besides testing, a more complex simulation uses various performance profiles to evaluate the long-term performance of students practicing real-life curricula. This simulation is used to verify that changes to the implementation do not introduce regressions in performance not caught by the test suite.

- The deliberate practice engine overrides the tutor's judgment and does not allow them to adapt to the student's needs.
  - The purpose of the engine is to create a stream of optimized exercises for the student, but it makes no claim to surface perfect exercises every time. Tutors have several tools to adapt the exercises more closely to the student's needs. They can skip exercises, mark courses or lessons as explicitly mastered so that they are never shown again, and use filters to select only exercises from certain tracks (for example, only showing comprehension and writing exercises).

- The user interface can introduce failure modes and additional training requirements.
  - The user interface has been made as sparse and simple as possible and is the easiest part of the intervention to train. The main screen, where students and tutors will spend the vast majority of their time, simply shows the examples in the current lesson. Tutors are entirely responsible for operating the software. They check off the examples the student read or spelled correctly, perform the intervention, and click on "Next Lesson"

to continue. All features of the interface could be explained in fifteen minutes because the complexity is hidden from users behind a single button.

- Literacy software might use tasks that do not transfer to real-world tasks, and students might only learn to perform well at the software tasks.
  - *Pictures Are For Babies* removes gamification, badges, games, multiple choice questions, and any other extraneous elements that might distract students from the core tasks.
  - Each of the tracks in the program has tasks identical to the real-world tasks they are designed to teach. Word courses teach students to read and spell words, sentence courses teach students to read and write sentences, and so on. The only difference is in the writing courses, where students are shown worked examples of the writing process. However, this limitation is addressed by having students complete writing production tasks outside the core software loop.
- Children will become bored and disengaged if the program is not filled with games, cartoons, or other engagement tricks.
  - The main antidote *Pictures Are For Babies* uses against boredom is not to pretend that learning to read and write is fun, but to engineer the conditions for flow and to design the intervention protocols to efficiently build the cognitive abilities needed for literacy.
  - The second antidote is to always pair the student with a tutor who can provide emotional support and encouragement much better than any gamification or speech recognition system could.
  - The reading break gives students a chance to rest while they orally develop their background knowledge, vocabulary, and comprehension skills. By giving them freedom to choose the books, they also develop their curiosity and intrinsic motivation to learn.
- Children are at various stages of developing their speech. Children also have varied dialects and accents that make a universal system difficult to implement.
  - Many programs try to automate the speech part of reading instruction with the use of speech recognition and AI. These systems are attractive to investors and some parents, but they introduce many failure modes. The decision to only rely on human tutors was also made to provide every student with a tutor of similar dialect and accent that can understand their speech without difficulty.
  - Tutors are also free to adapt the lessons to their specific dialects. For example, speakers of British English can change the spelling of words to match their dialect and use the

words on the screen as a guide to what words and patterns to teach, not as a rule of how to teach them.

- Deliberate practice is a demanding and effortful process that can be frustrating for students, and children need a more fun intervention.
  - It is true that deliberate practice is not inherently fun and the program has never tried to hide this fact. However, the other side of the coin is that deliberate practice produces a state of flow when the conditions are right. *Trane* attempts to engineer those conditions so that students can develop their intrinsic motivation to learn to read and write.
  - The reading break allows students to rest while developing their background knowledge and vocabulary. They are free to choose the books for this break to match their interests and curiosity.
  - Many of the programs in this field try to solve this problem by using gamification and cartoons. However, a meta-analysis on gamification shows that its effect size is only 0.49 and that the effect is smaller in longer interventions. While a decent strategy for short time interventions that care more about completion than mastery, the effects are not strong enough to replace a deliberate practice framework in a long-term literacy intervention.
  - A reason for picking the name *Pictures Are For Babies* is that these attitudes are offensive to children, who are naturally curious and who are blamed for the failures of their literacy education. The program takes full advantage of this curiosity and intrinsic motivation to learn, and teaches them the right cognitive abilities so that engagement tricks are not needed at all.
- Long term interventions will suffer from attrition and dropout, which will reduce the overall efficacy of the program.
  - Attrition is a problem for any long-term intervention, but *Pictures Are For Babies* is better positioned than most to mitigate it. Normal interventions are susceptible to attrition because even short breaks cause gaps in instruction that are difficult to recover from. By contrast, *Pictures Are For Babies* is resistant to small and medium breaks, and even to longer ones, because the student will continue from where they left off and the program will adjust to their new level of proficiency.
  - Permanent attrition is still a risk but mitigated by the full design of the program, from matching the material to the student's level, to the use of reading breaks, and to the

engineering of flow and intrinsic motivation.

  - Because the largest barriers to literacy occur in the early grades, struggling students that only complete the early stages of the program can have their trajectory permanently altered for the better. In the highly successful interventions Kilpatrick cites, students received only 20 to 80 hours of targeted remediation and their performance was even better when measured at yearly follow-ups. Once phonemic awareness deficits are remediated, students can continue to develop their sight vocabulary via the self-teaching hypothesis. Such dropouts would lose out on the full benefits of the program, but might still capture most of the benefits of these highly successful interventions.

- Writing production is outside the core loop of the program and requires additional effort from tutors.
  - While this is true, the decision to turn writing instruction into reading of worked examples brings the benefits of deliberate practice to writing instruction in a way programs based on production tasks cannot. The worked examples and increased volume of practice will provide students with a strong mental model for how writing works. Students that undertake the production part of the writing courses with the help of their tutors will transfer that model to real tasks.
  - The software and manual support tutors in selecting appropriate writing exercises and the worked examples help them design their own tasks and compare the student's work to the model answer. While this requires additional effort and cannot be tracked like the rest of the tasks, the program gives tutors the tools needed to conduct these exercises.

Further Evaluation

While *Pictures Are For Babies* has been designed to be a research-backed and effective literacy intervention and its claims are deeply rooted in well-researched and widely accepted principles, the depth of integration, the scope of the curriculum, and the ambition of its goals make further evaluation a critical next step. To that end, the author is engaged in an ongoing effort to find a research partner to conduct a multi-year longitudinal randomized controlled trial to evaluate the program's effectiveness in a real-world setting that can secure the necessary funding. Full access to the program will be provided at no cost to such a research partner and to all schools and students involved in the study. Due to the large efforts needed to study the full efficacy of the program, *Pictures Are For Babies* does not plan to conduct or coordinate any shorter-term studies that would

qualify for ESSA tiers 3 or 2.

The deep integration of Kilpatrick's orthographic mapping model in *Pictures Are For Babies* opens the possibility to also study its validity, providing the study with more scientific value than a mere product evaluation. Kilpatrick's model was used due to the strong results of the interventions he classified as highly successful and because it provided a clear instructional framework that an engineering effort cannot wait another decade to validate. The debate over the validity of this model concerns the extent to which phonemic proficiency is the missing ingredient that allows struggling readers to catch up to their peers. The opposing view states that segmenting and blending are sufficient for all students and that the ability to perform more complicated phoneme manipulations develops naturally as a byproduct of learning to read.

The proposed experimental design for the randomized controlled trial is as follows:

- A multi-year longitudinal study with a control group that receives its normal literacy instruction and two treatment groups, one using only the first tier of the word intervention (focused on only blending and segmenting) and the other using all three tiers of the word intervention (focused on blending, segmenting, and phoneme manipulation). Because of the scope of the program, the ideal study follows students from pre-K or first grade through the end of their K-12 education to capture its full effect.
- Participants are recruited from first grade cohorts in multiple schools, with random assignment to any of the groups. Students in the treatment groups will receive their instruction plus individual tutoring sessions at school and at home using *Pictures Are For Babies* for the duration of the study. Tutors for the treatment group that only uses the first tier will be instructed to not conduct any higher tiers even when the software issues a warning.
- All people assigned as tutors, both teachers and parents, will receive training on how to use the software and conduct the intervention protocols. Volunteers will be recruited to serve as tutors for students whose parents are unable to do so.
- Students will be evaluated at the end of each school year and at the end of the study using independent, standardized, norm-referenced assessments across the core areas of phonological processing, word-level reading, spelling, reading comprehension, and writing.
- Both treatment groups will be compared to the control group and to each other to determine the overall effect size of the program and the effect of the higher intervention tiers on the overall effect size. The comparison between the treatment groups will provide evidence on

whether explicit instruction on phoneme manipulation is necessary for struggling readers to catch up to their peers.

While a research partner is recruited, the author is collecting the full data from students whose parents agree to share it. Since *Pictures Are For Babies* runs locally and does not store any personal data, such sharing does not require any legal agreements or compliance infrastructure. The author is also looking to conduct school and individual free pilots that can provide feedback and iron out any issues in the program before a large-scale study is conducted. However, based on the extent and fidelity to which the research has been followed, the author is confident that any issues solved by user feedback are minor and will mostly affect the manual and other training materials, not the core of the program's design or software.

Annotated Bibliography

Books

The pedagogy outlined in *Pictures Are For Babies* is largely derived from the sources below. These are some of the most widely respected and research-backed works in their fields. Their authors are recognized leaders in reading science, writing instruction, and expert learning. These texts were not used for surface-level validation. They shaped the design and sequencing of every component in the program from the very beginning. Books were preferred over individual journal articles because they provide a more comprehensive and integrated view of the research that is difficult for an outside researcher to construct from individual articles.

- Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (Eds.). (2018). *The Cambridge Handbook of Expertise and Expert Performance* (2nd ed.). Cambridge University Press.
  - The definitive reference on deliberate practice and expert performance. Informed *Trane*'s design as a deliberate practice engine and the program's focus on mastery through effortful, structured training with immediate feedback.
- Hochman, J. C., & Wexler, N. (2024). *The Writing Revolution 2.0: A Guide to Advancing Thinking Through Writing in All Subjects and Grades*. Jossey-Bass.
  - The basis for the program's writing track. Informed the worked-examples approach, sentence-level to paragraph-level progression, and integration of writing with reading content.

- Kilpatrick, D. A. (2015). *Essentials of Assessing, Preventing, and Overcoming Reading Difficulties*. Wiley.
  - Synthesizes research on orthographic mapping, advanced phonemic awareness, and intervention effectiveness. Informed the tiered word-level intervention protocol and the categorization of reading interventions by standard score gains used in the Estimated Effect Size section.
- Snowling, M. J., Hulme, C., & Nation, K. (Eds.). (2022). *The Science of Reading: A Handbook* (2nd ed.). Wiley-Blackwell.
  - Comprehensive reference on the science of reading. Informed the program's adherence to research on phonological awareness, decoding, orthographic mapping, comprehension, and reading difficulties.

Scientific Studies

- 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.
  - Found one-to-one tutoring plus mastery learning produced an effect size of 2.0 on narrow tests. Mastery learning alone produced ~1.0. Originated the "2 sigma problem."
- Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, 145(11), 1029-1052. https://doi.org/10.1037/bul0000209
  - Meta-analysis of 59 studies (238 effect sizes). Overall interleaving effect g = 0.42. Stronger effects for complex material, longer retention intervals, and categories with high between-category similarity and within-category dissimilarity.
- Crissman, J. K. (2006). *The design and utilization of effective worked examples: A meta-analysis* [Doctoral dissertation, University of Nebraska-Lincoln]. UNL Digital Commons. https://digitalcommons.unl.edu/dissertations/AAI3208114
  - Meta-analysis of 62 studies on worked examples vs. problem-solving. Overall d = 0.57. Effect strongest for novices with low prior knowledge.
- Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, 104(4), 879-896. https://doi.org/10.1037/a0029185
  - Meta-analysis of 115 studies on writing interventions. Strategy instruction ES = 1.02;

SRSD = 1.17; text structure instruction = 0.59.

- Grissmer, D., White, T., Buddin, R., Berends, M., Willingham, D., DeCoster, J., Duran, C., Hulleman, C., Murrah, W., & Evans, T. (2023). A kindergarten lottery evaluation of Core Knowledge charter schools: Should building general knowledge have a central role in educational and social science research and policy? (EdWorkingPaper No. 23-755). Annenberg Institute at Brown University. https://doi.org/10.26300/nsbq-hb21
  - Lottery-based RCT of Core Knowledge curriculum in 9 charter schools (N = 2,310). Long-term treatment-on-treated effect on reading/ELA = 0.47 across grades 3-6.
- Kim, Y.-S. G., Wolters, A., & Lee, J. W. (2023). Reading and writing relations are not uniform: They differ by the linguistic grain size, developmental phase, and measurement. *Review of Educational Research*. https://doi.org/10.3102/00346543231178830
  - Meta-analysis of 395 studies (N = 120,669). Found reading comprehension and written composition are moderately correlated (r = .44).
- Kraft, M. A., Schueler, B. E., & Falken, G. (2026). What impacts should we expect from tutoring at scale? Exploring meta-analytic generalizability. *Review of Educational Research*. https://doi.org/10.3102/00346543261446660
  - Meta-analysis of 263 RCTs. Pooled effect size of 0.42 for all tutoring studies, dropping to 0.16-0.22 for large-scale programs (400+ students) using standardized tests.
- Kulik, J. A., Kulik, C.-L. C., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60(2), 265-299.
  - Meta-analysis of 108 controlled studies. Overall effect size of 0.52. LFM (Bloom's method) = 0.59. Follow-up exams (8 weeks later) = 0.71.
- Latimier, A., Peyre, H., & Ramus, F. (2020). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review*, 33(3), 959-987. https://doi.org/10.1007/s10648-020-09572-8
  - Meta-analysis of 29 studies on spaced retrieval practice. Overall g = 0.74 after correcting for publication bias. No significant difference between expanding and uniform spacing schedules.
- Mawson, R. D., & Kang, S. H. K. (2025). The distributed practice effect on classroom learning: A meta-analytic review of applied research. *Behavioral Sciences*, 15(6), 771. https://doi.org/10.3390/bs15060771
  - Meta-analysis of 22 classroom-based studies (31 effect sizes). Overall d = 0.54 for dis-

tributed vs. massed practice with curriculum-relevant materials.

- Moats, L. C. (2020). Teaching reading is rocket science: What expert teachers of reading should know and be able to do. *American Educator*, 44(2), 4-9, 39. https://www.aft.org/ae/summer2020/moats
  - Seminal article arguing that reading failure is preventable in all but a small percentage of children with serious learning disorders and that classroom instruction is the critical factor.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432-1463. https://doi.org/10.1037/a0037559
  - Meta-analysis of 61 studies on retrieval practice. Overall d = 0.50. Retrieval practice with feedback produced larger effects (d = 0.73) than without feedback (d = 0.39).
- Sailer, M., & Homner, L. (2019). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1), 77-112. https://doi.org/10.1007/s10648-019-09498-w
  - Meta-analysis of gamification on learning outcomes. Cognitive g = 0.49, motivational g = 0.36, behavioral g = 0.25. Shorter interventions had larger effects.
- Tipton, E., & Patton-Terry, N. (2026). We need an "engineering of reading": Why the "science of reading" may not be enough. *Mind, Brain, and Education*. https://doi.org/10.1111/mbe.70053
  - Argues that the science of reading requires an accompanying "engineering of reading" to translate research into effective classroom practice.
- Vroom, T.-A. (2021). *Adolescent writing instruction: A return to the sentence* [Doctoral dissertation, St. John's University]. St. John's Scholar. https://scholar.stjohns.edu/theses_dissertations/213/
  - Qualitative study with an embedded quantitative analysis of The Writing Revolution's sentence-level strategies. d = 0.84 for writing quality after 4 months of instruction with 9th graders. Not peer-reviewed.

Literacy Assessments

The data on the state of literacy instruction cited in the problem statement are drawn from the following three assessments. Together they cover school-age students (NAEP, PISA) and adults (PIAAC), providing a complete picture of literacy outcomes across grades and countries.

- National Center for Education Statistics. (2024). *NAEP Report Card: Reading — National trends and student skills, grade 12*. U.S. Department of Education.

https://www.nationsreportcard.gov/reports/reading/2024/g12/national-trends/

  - 5% of 12th graders scored at the NAEP Advanced level in 2024. 32% scored below Basic.

- OECD. (2023). *PISA 2022 Results (Volume I): The state of learning and equity in education.* OECD Publishing. https://ilsa-gateway.org/studies/results/1612

  - 7% of 15-year-olds scored at the highest proficiency levels (5 or 6) across OECD countries. 26% scored below Level 2.

- OECD. (2024). *Survey of Adult Skills 2023: United States — Country note.* OECD Publishing. https://www.oecd.org/en/publications/survey-of-adults-skills-2023-country-notes_ab4f6b8c-en/united-states_427d6aac-en.html

  - 13% of US adults scored at the highest levels (4 or 5) in literacy. 28% scored at Level 1 or below.