

PORTFOLIO CASE STUDY

Assessment-to-Instruction Cycle (AIC):
Finding the Instructional Boundary

Designing an Evidence-Guided Assessment-to-Instruction System for a Community Literacy Cohort

Project Overview

For a 12-student reading cohort spanning Grades 2–9, I developed and implemented a structured process for translating broad assessment results into specific, testable instructional decisions.

The central problem was straightforward but consequential: standardized assessment data could identify areas of relative strength and weakness, but often did not identify where instruction should actually begin.

My role was to build the translation layer between assessment and instruction.

Rather than treating a low score as an instructional prescription, I developed a system that moved from:



The system was deliberately iterative. Each new source of evidence could refine an earlier conclusion, but it could not silently rewrite the evidence that had originally supported it.

Cohort and Baseline Context

The Fall 2026 cohort contained 12 students across Grades 2–9, with a median enrolled grade of 5. The corrected source-control process preserved verified program grade separately from conflicting assessment metadata rather than silently reconciling discrepancies. It also explicitly treated floor scores, ceiling scores, and “not tested” results differently.

I recalculated cohort descriptives from the Fall baseline raw-score export using those verified grades.

Descriptive indicator	Cohort result
Students	12
Grade span	2–9
Median grade	5
Word Recognition placement indicator below enrolled grade	7/12 — 58%
Reading Comprehension placement indicator below enrolled grade	6/11 tested — 55%
Oral Vocabulary placement indicator at or above enrolled grade	8/12 — 67%
Broad DORA summary/WS indicator below enrolled grade	8/12 — 67%
Reported Lexile point estimate below 420	6/12 — 50%
Median reported Lexile point estimate	approximately 400

These comparisons are intentionally descriptive. DORA grade-equivalent values were treated as placement indicators rather than equal-interval measurements, instructional ages, or exact quantities of delay. Broad total scores were also not used to select instructional targets.

That distinction mattered. The cohort contained substantial within-student variation that could disappear inside an overall reading score.

For example, the reconciliation process identified a recurring pattern in which oral/verbal knowledge was substantially stronger than access to print. Five students were specifically flagged for this pattern. At the cohort level, the dominant unresolved issues were not simply “low reading”: they included print-access bottlenecks, decoding-to-encoding discrepancies, familiar-word versus novel-word differences, and questions about whether comprehension difficulty was primarily constrained by reading or remained present when language was presented orally.

The Assessment-to-Instruction Architecture

The original assessment-development work evolved into a 12-phase analytic architecture:

1	Source Audit — establish which records were current, historical, missing, invalid, duplicated, contradictory, floor-censored, or ceiling-censored.
2	Measurement Map — specify what each instrument could and could not support.
3	DORA-Only Model Freeze — interpret direct academic evidence before looking at cognitive data.
4	MindPrint-Only Model Freeze — interpret cognitive performance independently of academic scores.
5	Scientific Evidence Crosswalk — constrain interpretation using instrument documentation and relevant research.
6	DORA × MindPrint Reconciliation — compare the independent models without allowing one instrument to overwrite the other.
7	Discriminating Probe Design — identify the smallest additional observation capable of changing an instructional decision.
8	Instructional Leverage Analysis — identify the highest-payoff currently accessible skill.
9	Formative Monitoring Selection — select a repeatable short measure aligned to that instructional boundary.
10	Tutor Translation — convert the analytic model into usable instructional directions.
11	Longitudinal Updating — revise the model as probe, formative, session, and transfer evidence accumulated.
12	Cohort/Program Evaluation — examine recurring instructional patterns without erasing individual variation.

DORA-only model freeze Direct academic evidence	MindPrint-only model freeze Cognitive hypotheses
DORA × MindPrint reconciliation Compare the independent models without allowing one source to overwrite the other.	

A particularly important design decision was divergence before convergence.

DORA and MindPrint were not immediately combined into a single student “profile.”

The DORA model was frozen first. The MindPrint model was then developed independently. Only after both existed were they reconciled.

MindPrint was explicitly prevented from directly selecting an academic target. Its appropriate role was to generate hypotheses about instructional conditions—such as load, retention, pace, or output demands—that could then be tested against direct academic performance.

This prevented an appealing cognitive explanation from replacing evidence about what the student could actually read, decode, spell, or comprehend.

From Assessment Discrepancy to Diagnostic Question

For each student, I converted assessment patterns into competing explanations rather than immediate conclusions.

A typical reasoning sequence looked like this:

OBSERVATION	Familiar-word recognition is much stronger than spelling.
POSSIBLE EXPLANATIONS	<p>The student may have an encoding/orthographic weakness.</p> <p>The apparent difference may reflect handwriting or output demands.</p> <p>The result may be an artifact of adaptive item sampling.</p> <p>Retention may be contributing.</p> <p>The student may know practiced words but have difficulty generalizing patterns.</p>
DECISION QUESTION	What is the smallest observation that would distinguish among explanations that would actually lead to different instruction?

This became the core design principle of the probe system:

Do not collect more data because more data are available. Collect the smallest amount of new evidence that could change the instructional decision.

The field kit formalized that rule and required convergence across multiple observations rather than reliance on a single cutoff. It preserved first responses, final responses, prompts, interruptions, self-corrections, and validity conditions. Four consecutive errors triggered a stopping rule rather than converting unadministered items to failures.

The probe system therefore functioned less like another achievement test and more like a set of brief instructional experiments.

Probe Engineering

The assessment work produced a reusable library of discriminating contrasts.

Alphabetic Transfer Ladder

This route tested whether phonemic knowledge transferred progressively into print:

oral sound work → phoneme-to-letter mapping → familiar real-word reading → pseudoword decoding → spelling

Its purpose was to find the first point at which otherwise available knowledge became unstable.

Word-Form Precision and Encoding

The WFPE contrasted real-word recognition, invented-word decoding, and spelling using structurally matched items.

Its decision question was:

Does known-word recognition exceed generative decoding or exact spelling?

The forms were deliberately treated as locally constructed instructional probes rather than psychometrically equivalent tests.

Output-Mode Contrast

The OMC compared written spelling with oral or movable-letter spelling while holding content as constant as possible.

This was used only when the result could determine whether reduced handwriting demand was a meaningful instructional support. A difference between modes was explicitly treated as a moderator hypothesis—not proof of a motor deficit.

Listening-versus-Reading Comprehension

The LRC compared comprehension of equivalent but different passages presented through listening and reading.

Its purpose was to distinguish:

difficulty accessing meaning through print / difficulty that remained when print demands were removed.

Orthographic Retention

The ORT was used selectively when retention could actually change instructional spacing or review. Students learned unfamiliar printed forms and were checked immediately and again at the following session. It was explicitly not treated as a general memory test.

Most importantly, students did not receive a standard cohort battery. Each student was routed only through modules capable of resolving their own decision question.

Implementation Engineering

The analytical design was only useful if tutors could administer it consistently.

I therefore translated the diagnostic system into individual student probe packets rather than asking tutors to interpret a master assessment battery.

Each packet was engineered so that the correct next action was obvious:

what to print → what materials to prepare → exactly what to say → what to show the student → what to record → when to stop → what not to interpret

Tutors were not asked to determine a student's diagnostic route, calculate scores across modules, choose a new instructional target, or interpret psychometric meaning.

Their role was to produce high-quality direct observations.

My role was to interpret those observations in combination with the assessment record.

This division of labor reduced administration burden while protecting the quality of the resulting evidence.

Selecting the Formative Measure

After identifying the likely instructional boundary, I selected a short formative assessment aligned to the highest-leverage currently accessible reading skill.

The formative was not meant to rediscover the entire reading profile. It answered a narrower question:

Is performance on the skill we are actually teaching becoming more independent, accurate, and transferable?

≥90% consider advancing	Clear prerequisite failure move downward	Otherwise hold and continue targeted instruction
-----------------------------------	--	--

Importantly, formative results could contradict the original placement hypothesis.

When they did, the target changed.

That made progress monitoring a decision system rather than a compliance exercise.

De-Identified Examples of the Reasoning Process

Case A — Broad “Foundational Reading” Concern Became a Specific Transfer Problem

An elementary student entered the process with DORA word recognition and spelling at the apparent reporting floor, low phonics, mixed phonemic-awareness performance, and very strong oral vocabulary.

A superficial interpretation could easily have resulted in broad beginning-reading remediation.

The targeted probe changed that model.

The student completed oral sound work 4/4, familiar real-word reading 8/8, and simple spelling 6/8. Performance weakened specifically when the student had to apply known vowel patterns to unfamiliar words: pseudoword reading fell to 5/8.

A subsequent Long Vowel formative was 5/10, with plausible but inaccurate substitutions such as reading one long-vowel pattern as another.

The conclusion changed from:

FROM	“broad foundational decoding weakness”
TO	“available basic alphabetic knowledge with unstable generative long-vowel transfer and encoding.”

Instruction consequently concentrated on contrastive silent-e work, real-word-to-novel-word transfer, spelling/word building, and checking every part of the printed word.

The probe prevented unnecessary reteaching of skills the student already possessed.

Case B — Reading Was Stronger Than Spelling, but the Question Was Why

Another elementary student had stronger familiar-word reading and comprehension than phonics and spelling.

The baseline evidence supported several reasonable hypotheses:

- a broad decoding problem,
- a more specific spelling/word-form problem,
- or an effect of slower written output.

A fixed probe produced 7/8 real-word reading and 6/8 invented-word reading, while the spelling block reached the pre-established four-error stop rule. Errors preserved much of the sound structure but selected inaccurate conventional spellings.

The subsequent Diphthongs formative was 7/10, with errors among competing vowel representations.

The combined evidence moved the model away from broad decoding and toward less-stable orthographic retrieval and selection, particularly when the student had to generate the spelling rather than recognize a word.

The resulting activities contrasted plausible vowel spellings, required independent retrieval before feedback, and used a spell → compare → repair → reread cycle.

Case C — A 100% Formative Prevented Unnecessary Remediation

An adolescent entered the project with DORA showing strong oral vocabulary and phonemic awareness relative to much weaker print-based scores.

The probe found inaccurate first-pass decoding, stronger performance on familiar than unfamiliar forms, frequent self-correction, and reading that exceeded the probe's four-minute completion window.

A reasonable hypothesis was that common vowel-team knowledge itself was weak.

The subsequent Vowel Digraphs formative answered that question decisively:

10/10, including novel items.

The low-level skill was available under direct structured assessment and therefore should not be retaught as though absent.

The revised instructional model became:

known phonics knowledge is present; the more meaningful boundary is efficient, complete application of that knowledge across unfamiliar, increasingly complex words and connected text.

Tutor session evidence reinforced the distinction: elementary-level activities were frequently too easy, while authentic text exposed slow decoding, partial-cue substitutions, longer-word difficulty, and spelling weakness.

Instruction therefore moved upward toward full-word analysis, multisyllabic decoding, decoding–spelling reciprocity, and connected-text transfer.

This case demonstrates one of the system's most important functions:

assessment is allowed to rule out instruction as well as prescribe it.

Using Tutor Session Notes as Evidence

Session reports were not treated merely as proof that tutoring occurred.

I reviewed them for repeated behavioral and instructional signals such as:

first response versus self-correction, familiar versus unfamiliar word performance, spontaneous strategy use, prompts required, transfer into connected reading, recurring substitution patterns, persistence, task conditions, and changes after explicit instruction.

This allowed session evidence to serve as an ecological check on formal assessment.

A standardized measure might reveal a discrepancy.

A probe might locate the likely boundary.

A formative might establish whether a target was currently mastered.

But the tutor's observations often showed how the problem manifested during authentic learning.

When the same pattern appeared across assessment, probe, formative, and ordinary tutoring, confidence in the instructional model increased.

When the sources disagreed, the disagreement became the next question rather than something to average away.

Progress Review and Instructional Translation

The final synthesis for each student combined:

baseline assessment → probe rationale and result → formative performance → tutor-session evidence → updated instructional model → next progress-monitoring decision → activity/lesson packet

The resulting progress reviews were designed to answer two different professional needs simultaneously.

FOR THE ANALYST	What does the accumulated evidence currently support, what remains uncertain, and what evidence would change the model?
FOR THE TUTOR	What should I actually do with this student at the next session?

That distinction led to the development of short tutor-facing activity packets built around the strongest current instructional target.

Activities were selected not because they were generically engaging, but because they recreated the specific learning contrast revealed in the student's evidence—for example:

familiar word → unfamiliar transfer word
read → spell → read again
short vowel → silent-e transformation
first attempt → compare → repair
known word pattern → novel word using the same pattern

The objective was to make instructional activity itself produce additional useful evidence.

What the Project Produced

The work generated a connected set of professional artifacts:

1	source-controlled assessment audit
2	DORA-only analytic model
3	MindPrint-only analytic model
4	scientific evidence crosswalk
5	DORA/MindPrint reconciliation
6	discriminating probe system
7	individual tutor probe packets
8	formative assignments
9	progress-review syntheses
10	evidence-based activity banks
11	direct mini-lesson packets

Together, these artifacts document not simply student performance, but the evolution of instructional decisions.

Professional Competencies Demonstrated

Competency	Evidence from the project
Assessment literacy	Distinguished direct findings, derived patterns, missing data, invalid results, ceilings/floors, and hypotheses
Diagnostic reasoning	Generated competing explanations rather than treating scores as causes
Evidence integration	Triangulated standardized assessment, cognitive data, probes, formatives, and authentic session observations
Progress monitoring	Selected short measures tied to the current instructional boundary
Instructional design	Converted error patterns into targeted, transferable reading activities
Experimental thinking	Used matched contrasts, alternate forms, stopping rules, and replication where decisions depended on uncertain results
Data quality / epistemic discipline	Preserved contradictory and invalid evidence rather than smoothing it away
Tutor enablement	Designed administration packets and lesson routines that minimized unnecessary tutor decision-making
Differentiation	Changed both instructional target and instructional conditions according to individual evidence
Program design	Created a repeatable architecture capable of operating across an entire heterogeneous cohort

Portfolio Framing

The central professional contribution of this project was not the production of 12 progress reports.

It was the creation of a repeatable instructional decision architecture for determining what broad assessment results actually meant for the next act of teaching.

The system was designed around one recurring question:

Where is the student's current instructional boundary—and what is the smallest amount of evidence needed to locate it responsibly?

That question changed the role of assessment.

Assessment stopped being the endpoint of evaluation and became the beginning of a disciplined cycle of hypothesis, observation, instruction, monitoring, and revision.

Confidentiality Note

All public portfolio versions should remove student names, family information, schools, tutor identities, birthdates, assessment IDs, and exact service dates. Case examples should be labeled by broad developmental band rather than exact identifying combinations wherever possible. Cohort statistics should remain aggregated.

AI Collaboration Disclosure

This project was developed through an educator-directed collaboration with OpenAI's ChatGPT. I defined the instructional problem, selected and contextualized the evidence sources, established practical and pedagogical constraints, reviewed the resulting analyses, corrected source and interpretation errors, and retained responsibility for instructional decisions.

The AI served as an analytic and design collaborator rather than solely as a writing tool. Its contributions included helping structure independent assessment analyses, generate and compare competing hypotheses, design targeted discriminating probes and decision rules, synthesize assessment and tutoring evidence, calculate descriptive summaries, review relevant research, and translate findings into progress-monitoring and instructional materials.

Several elements of the final methodology therefore emerged iteratively through human–AI co-design. AI-generated analyses and materials were treated as provisional and subject to review against source data and direct instructional evidence. The AI did not assess or observe students, administer measures, supervise instruction, or independently determine student services or educational decisions.