



CONSTRUCTING RESPONSIVE AN EVALUATION MODEL FOR EARLY READING BASED ON FORMATIVE DIAGNOSTIC ASSESSMENT

Syamzah Ayuningrum¹, Endry Borieswati², Aceng Rahmad³

Universitas Negeri Jakarta^{1,2,3}

Email: syamzah_9906925026@mhs.unj.ac.id¹, endry.boeriswati@unj.ac.id²,

acengrahmat@unj.ac.id³

ABSTRACT

This study aims to construct a Responsive Diagnostic Evaluation Model (RDEM) for early reading assessment based on formative diagnostic assessment. The study employed a conceptual model-development approach through the stages of problem identification, theoretical synthesis, construct formulation, instrument mapping, development of decision-making rules, and internal model evaluation. The model was developed by integrating Stake's responsive evaluation, diagnostic assessment, formative assessment, differentiated instruction, and data-informed feedback. The findings produced five interconnected domains: responsive context, diagnostic antecedents, student-profile classification, feedback and differentiated instruction, and responsive outcomes. These domains operate within an adaptive cycle that connects students' initial conditions, the identification of specific reading difficulties, instructional decision-making, follow-up actions, progress monitoring, and meaningful outcomes. The model also provides decision rules that translate diagnostic evidence into targeted instructional actions based on students' learning profiles. RDEM positions evaluation as a responsive and continuous instructional control system rather than merely a mechanism for measuring and reporting scores. As the model remains conceptual, further expert validation and classroom-based testing are required to examine its feasibility and effectiveness.

Keywords: *responsive evaluation; formative diagnostic assessment; early reading; differentiated instruction; feedback*

INTRODUCTION

Early reading is the foundation through which children gain access to almost every subsequent school activity. In Phase A of primary education, students are expected to develop phonological awareness, recognize letters and their corresponding sounds, combine syllables, decode words, read short texts with increasing fluency, and understand simple information. These abilities do not develop as one undifferentiated skill. They form an interconnected system in which weakness in one component can constrain the development of the others. Ehri (2005), for example, explains that word-reading development depends on increasingly stable connections among graphemes, phonemes, pronunciations, and meanings. When these connections remain fragile, students may appear to have the same general reading problem even though the underlying sources of difficulty are different.

The consequences of weak early reading are not confined to language lessons. Students who read inaccurately or laboriously obtain less effective practice from printed materials and expend more cognitive effort on identifying words, leaving fewer resources for understanding text. Stanovich (1986) described this cumulative pattern through the

Matthew effect: 'early differences in reading experience can widen over time as stronger readers gain more practice while struggling readers encounter repeated failure. Early identification is therefore not an administrative exercise. It is a preventive instructional requirement because delayed or imprecise identification allows a manageable difficulty to become a broader barrier to learning.

In many classrooms, however, assessment of early reading still functions as a general performance check. Teachers ask students to name letters, read words, or read a short passage and then classify them as fluent, less fluent, or unable to read. Such labels describe the visible outcome but do not explain whether the difficulty originates in phonological awareness, letter-sound correspondence, syllable blending, word decoding, automaticity, vocabulary, or simple comprehension. A student who cannot read a word because the letter sounds are unstable requires a different response from a student who knows the sounds but cannot blend them. Treating both students as merely 'weak readers' removes the information needed for precise intervention.

Evidence from previous research shows that this problem is not limited to one classroom. Schmitterer & Brod (2021) found that elementary teachers' judgments about students' need for reading intervention were often based on informal experience and selected indicators rather than a systematic combination of standardized reading-related data. Their findings indicate that teacher judgment is indispensable but can be biased or incomplete when it is not supported by structured diagnostic evidence. These studies confirm that the practical challenge is not simply whether teachers assess, but whether the assessment evidence is sufficiently specific and consistently interpreted.

Screening and progress measures offer an important response, yet they do not automatically produce diagnostic understanding. Amendum et al. (2021) argued that oral reading fluency is useful for screening but should be followed by more diagnostic measures to explain students' reading performance. Cognitive diagnostic studies similarly demonstrate the value of decomposing reading performance into fine-grained attributes. Toprak & Çakır (2021) developed a cognitive diagnostic reading test to identify patterns of mastery rather than report only an overall score, while international analysis of fourth-grade reading data has shown that cognitive diagnostic approaches can generate detailed student profiles from large-scale reading assessments. The contribution of these studies is clear: assessment becomes more instructionally meaningful when it reveals a structure of strengths and weaknesses.

Nevertheless, fine-grained diagnosis alone does not guarantee improved instruction. Peters et al. (2022) examined learning-progress assessment and assessment-based differentiated reading materials in second-grade classrooms. The intervention did not produce higher reading gains than business-as-usual instruction. This result is analytically important because it challenges the assumption that providing more data or materials automatically changes learning. Assessment evidence must be interpreted accurately, translated into explicit instructional decisions, implemented with sufficient quality, and reviewed against subsequent student responses. Without this evidence-to-action chain, progress monitoring can become another reporting routine rather than a mechanism for instructional adaptation.

Formative assessment provides the principle required to establish this connection. Black & Wiliam (1998) emphasize that assessment is formative when evidence about



learning is used to modify teaching and learning activities. Wiliam (2011) further explains that formative practice clarifies where learners are, where they are going, and how they can move forward. In early reading, this means that the value of an assessment is not determined by the number of tasks administered but by whether the resulting evidence changes the next instructional step. Feedback must therefore be specific enough to identify the problem, actionable enough to guide practice, and sufficiently continuous to show whether the chosen response is working.

Diagnostic assessment strengthens this formative function by separating broad performance into interpretable components. Cognitive diagnostic assessment has been developed precisely to provide fine-grained information about latent knowledge structures and patterns of attribute mastery beyond a single score (Mei & Chen, 2024). However, much of this work has focused on psychometric modelling, large-scale tests, or second-language reading comprehension. It offers strong procedures for identifying profiles, but it does not by itself provide a complete classroom evaluation framework that incorporates students' initial context, teacher concerns, instructional transactions, and the meaning of outcomes in everyday early-grade learning.

Stake's responsive evaluation offers the contextual dimension missing from narrowly test-centred approaches. Responsive evaluation directs attention to antecedents, transactions, outcomes, and the issues considered important by stakeholders (Stake, 1975, 2004). In early reading, antecedents include students' prior literacy experiences and initial abilities; transactions include teacher-student interaction, task design, scaffolding, and students' responses; and outcomes include not only score changes but also increased independence, participation, and confidence. Yet responsive evaluation is primarily designed to understand and judge educational activity. Without operational diagnostic indicators and explicit decision rules, it can remain rich in description while offering limited guidance about which reading component should be addressed next.

Differentiated instruction supplies the instructional orientation needed after diagnosis. (Tomlinson, 2001, 2014) argues that instruction should be adapted to learner readiness and need through changes in content, process, task, grouping, and support. The limitation is that differentiation is only as accurate as the evidence on which it is based. If students are grouped through broad impressions, differentiation may reproduce the same imprecision as conventional instruction. A valid model must therefore connect component-based diagnosis to dynamic profiles, and then connect those profiles to targeted actions that can be revised when progress evidence contradicts the initial interpretation.

The initial problem analysis at SDIT Al Amal Bekasi reflects this wider conceptual gap. Teachers recognized heterogeneous early-reading readiness, yet assessment was not consistently organized through explicit indicators. Students were commonly asked to read letters, syllables, words, or short texts, but recurring errors were not always analysed to determine their probable source. The results consequently produced broad descriptions rather than differentiated instructional priorities. Home-literacy exposure, confidence, participation, and classroom response were also not systematically integrated into the interpretation. The problem was therefore not the total absence of assessment, but the

fragmentation of context analysis, diagnosis, feedback, differentiation, and outcome evaluation.

Previous research has developed each of these elements separately. Responsive evaluation provides contextual sensitivity; cognitive diagnostic assessment generates detailed profiles; formative assessment connects evidence with feedback; progress monitoring supplies repeated data; and differentiated instruction adapts teaching to learner needs. The unresolved gap lies in the absence of a unified and operational model that connects these strengths in a continuous early-reading evaluation cycle. Existing studies commonly stop at one of four points: identifying risk, producing a diagnostic profile, delivering feedback, or testing a differentiated intervention. They do not consistently specify how contextual information and diagnostic evidence are converted through decision rules into instruction, how the response is monitored, and how outcomes are judged against the original source of difficulty.

To address this gap, this study constructs the Responsive Diagnostic Evaluation Model (RDEM) for early reading. RDEM integrates Stake's responsive evaluation with formative diagnostic assessment, dynamic student-profile classification, differentiated instruction, progress monitoring, and responsive outcomes. Its central proposition is that evaluation should function as an evidence-to-action system: contextual evidence frames the diagnosis; diagnostic evidence produces a provisional profile; the profile activates an explicit decision rule; the decision guides targeted instruction; and outcome evidence confirms, revises, or rejects the initial interpretation. Accordingly, the study asks: How can a responsive evaluation model for early reading be constructed by integrating contextual analysis, formative diagnostic assessment, student-profile classification, instructional decision rules, and responsive outcomes? The study aims to formulate the conceptual structure, operational domains, and implementation flow of RDEM and to clarify its theoretical contribution and limitations.

METHOD

This study employed a conceptual model-development design to construct a theoretically coherent and operational framework for evaluating early reading. The study did not test the effectiveness of RDEM or claim empirical improvement in students' reading outcomes. The unit of analysis was the relationship among the problems identified in early-reading assessment, the theoretical principles used to explain those problems, and the components required to convert assessment evidence into instructional decisions. The empirical context described in the group presentation at SDIT Al Amal Bekasi was used as the practical point of departure, particularly the absence of systematic initial identification, unclear diagnostic indicators, broad student classification, uniform instruction, and limited progress monitoring.

Model construction was conducted through four integrated procedures. First, the practical problems were reduced into core evaluative needs: contextual identification, component-based diagnosis, profile classification, actionable feedback, differentiated instruction, and outcome interpretation. Second, a theoretical synthesis was undertaken using Stake's responsive evaluation, Black and Wiliam's formative assessment, Wiliam's assessment-for-learning principles, diagnostic reading assessment, and Tomlinson's differentiated instruction. Third, the synthesis was translated into model domains,



subconstructs, operational indicators, expected outputs, and illustrative decision rules. Fourth, the internal coherence of the model was examined by tracing whether each type of evidence had a defined interpretive function and whether every diagnostic conclusion led to a plausible instructional response and a corresponding progress indicator.

The main data consisted of the problem analysis, theoretical matrix, proposed diagnostic instrument, implementation strategy, and model diagram contained in the source presentation, supplemented by relevant scholarly literature used to clarify the research gap and strengthen the theoretical relations among constructs. Analysis was conducted through conceptual coding, comparison, and synthesis. Concepts with similar functions were grouped, overlapping components were consolidated, and causal or operational relations were mapped into an adaptive evaluation sequence. The resulting model was evaluated conceptually according to clarity of constructs, consistency between indicators and outputs, continuity from diagnosis to instruction, and the capacity of outcome evidence to revise earlier decisions. Because expert validation, classroom trials, and psychometric testing were not conducted, the findings are presented as a conceptual model requiring subsequent empirical validation.

RESULTS AND DISCUSSION

1. Problems and Conceptual Gap

The gap is sharper than the simple claim that teachers need more assessment. Adding another test would not solve the problem if the test still ends in a score. What is missing is an evaluative architecture that preserves contextual understanding while imposing operational discipline. Such an architecture must answer five questions in sequence: What conditions shape the student's performance? Which reading component is weak? What profile best represents the current need? Which instructional action logically follows? Did the resulting change occur in the component that was targeted? RDEM was designed around this sequence.

These weaknesses reinforce one another. When the construct is vague, the instrument samples reading behavior without determining the source of error. When the source is unknown, student grouping is based on overall performance rather than need. When grouping is imprecise, the same activity is delivered to students with different difficulties. When intervention is not aligned with the diagnosis, an unchanged score cannot reveal whether the student failed to respond, the strategy was inappropriate, or the original diagnosis was incorrect. The practical problem is therefore a broken evidence chain: observation does not consistently become diagnosis, diagnosis does not reliably become instruction, and outcome does not systematically revise the next decision.

The analysis indicates that the observable classroom problem is produced by a chain of weaknesses rather than by one isolated deficiency. The first weakness occurs before instruction: no structured baseline is established. The second concerns construct clarity: reading is treated as a single ability even though the same incorrect response can originate from different phonological, orthographic, fluency, or comprehension processes. The third concerns data use: results are recorded as broad classifications but are not converted into instructional priorities. The fourth concerns continuity: progress is checked through

ordinary classroom tests rather than through repeated measures of the component originally identified as weak.

2. Construct of the Responsive Diagnostic Evaluation Model

The theoretical synthesis produced five interrelated domains: responsive context, diagnostic antecedents, student-profile classification, feedback and differentiated instruction, and responsive outcomes. Table 1 presents the construct, operational indicators, expected outputs, and theoretical sources of each component.

Table 1. Construct of the Responsive Diagnostic Evaluation Model (RDEM)

No.	Model Domain	Subconstruct	Operational Indicators	Expected Output	Source
1	Responsive Context	Identification of needs and contextual conditions	Identify classroom reading problems, stakeholder concerns, instructional conditions, and home literacy background.	Contextual profile of early reading instruction	Stake (1975, 2004)
2	Diagnostic Antecedents	Initial reading readiness	Assess phonological awareness, letter recognition, grapheme-phoneme correspondence, syllable blending, decoding, fluency, and basic comprehension.	Initial reading-ability map	Black & Wiliam (1998); Wiliam (2011); Ehri (2005)
3	Diagnostic Antecedents	Specific diagnosis of difficulty	Record omission, substitution, reversal, hesitation, repetition, blending, fluency, and comprehension errors.	Individual diagnostic profile	Ehri (2005); Toprak & Çakır (2021)
4	Student-Profile Classification	Classification of learning needs	Group students by dominant difficulty, mixed profile, priority need, and current readiness.	Dynamic student-profile classification	Min & He (2022); Tomlinson (2014)
5	Feedback and Differentiated Instruction	Instructional decision rules	Translate diagnostic evidence into targeted activities, flexible groups,	Differentiated instructional plan	Tomlinson (2001, 2014); Wiliam



No.	Model Domain	Subconstruct	Operational Indicators	Expected Output	Source
			scaffolding, and task adjustment.		(2011)
6	Feedback and Differentiated Instruction	Progress monitoring	Provide actionable feedback, reassess target skills, compare baseline and progress evidence, and revise intervention.	Progress-monitoring record and instructional decision	Black & Wiliam (1998); Tang & Zhan (2021)
7	Responsive Outcomes	Specific and contextual outcomes	Identify changes in reading skill, independence, engagement, confidence, and alignment with initial needs.	Meaningful interpretation of outcomes	Stake (1975, 2004)
8	Continuous Adaptive Cycle	Evaluation-based improvement	Use outcome evidence for re-diagnosis, profile revision, and further instructional adjustment.	Sustainable evaluation cycle	Stake (2004); Wiliam (2011)

The construct indicates that RDEM does not separate evaluation from instruction. Responsive context prevents test results from being interpreted without regard to learning conditions. Diagnostic antecedents identify the component and pattern of difficulty. Student-profile classification condenses multiple observations into a temporary instructional hypothesis, while feedback and differentiated instruction test that hypothesis through targeted action. Responsive outcomes then determine whether the observed change corresponds to the original diagnosis. The model is therefore not a sequence of administrative stages but a system of linked inferences: each stage must be justified by evidence from the preceding stage and remain open to revision by evidence from the following stage.

The model also separates evidence from decision. A reading error is evidence; the interpretation that it reflects weak phonological integration is a diagnostic inference; the choice to provide guided blending practice is an instructional decision. Conflating these stages encourages premature conclusions. By making the stages visible, RDEM requires teachers to justify why a particular intervention follows from a particular pattern and to revise that inference when progress evidence does not support it.

The distinction between a profile and a label is critical. A label such as “weak reader” appears stable and global, whereas a profile is provisional, component-based, and

instructionally useful. A student may initially show a dominant syllable-blending difficulty, improve after targeted practice, and subsequently require fluency support. RDEM therefore treats classification as dynamic. This design reduces the risk that assessment fixes students into permanent ability groups and instead uses grouping as a temporary response to current evidence.

3. Flow of RDEM

Figure 1 presents a simplified flow of the model. The five domains form an adaptive cycle rather than a one-way sequence. Outcome evidence returns to the contextual and diagnostic stages so that the student profile and instructional response can be updated.

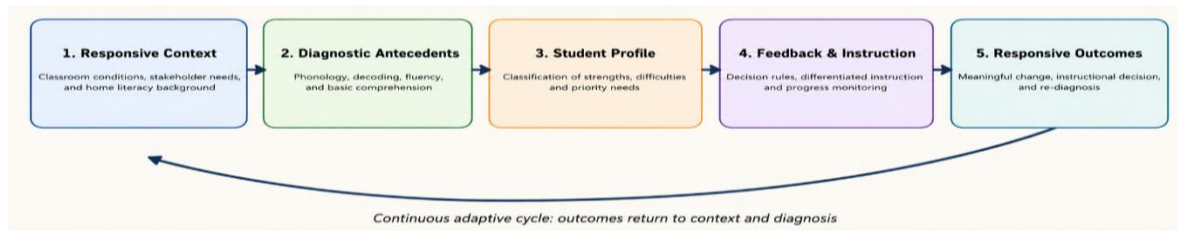


Figure 1. Simplified flow of the Responsive Diagnostic Evaluation Model for early reading

4. Diagnostic Domains of Early Reading

RDEM organizes diagnostic evidence into phonological ability, decoding and fluency, and basic comprehension. These domains distinguish the source of difficulty rather than reducing reading performance to a single score.

Table 2. Diagnostic domains and indicators of early reading

Domain	Indicators	Suggested Assessment	Diagnostic Meaning	Source
Phonological ability	Sound awareness; syllable segmentation; syllable blending	Oral sound and syllable tasks	Difficulty indicates weak analysis or integration of speech-sound units.	Ehri (2005)
Decoding	Letter recognition; letter-sound correspondence; syllable and word reading	Letter cards, syllable lists, and graded word lists	Errors reveal unstable recognition, correspondence, or blending.	Ehri (2005)
Fluency	Accuracy, continuity, hesitation, repetition, and dependence on	Oral reading of a short text	Accurate but slow reading indicates limited automaticity.	Stanovich (1986)



Basic comprehension	prompting Explicit information, simple inference, and retelling	Questions and oral retelling based on a short text	Difficulty indicates weak retrieval, integration, or retention of information.	Toprak & Çakır (2021); Khamman & Meejang (2024)
----------------------------	--	--	--	---

5. Decision Rules and Instructional Adaptation

The model's operational contribution lies in its decision rules. A diagnostic result is not sufficient unless it leads to a specified instructional response. Table 3 illustrates the relationship between findings, interpretations, actions, and progress indicators.

Table 3. Illustrative decision rules for early reading instruction

Diagnostic Finding	Interpretation	Instructional Decision	Progress Indicator
Cannot identify or distinguish sounds consistently	Weak phonological awareness	Use sound matching, initial-sound identification, and oral word play.	Target sounds are identified more consistently.
Recognizes letters but cannot blend syllables	Weak phonological integration	Provide modelled and guided syllable-blending practice.	Syllables are combined with less prompting.
Reads accurately but very slowly	Weak automaticity	Use repeated reading, phrase reading, and controlled fluency practice.	Reading becomes more continuous with fewer hesitations.
Reads fluently but cannot answer explicit questions	Weak vocabulary or literal comprehension	Teach key vocabulary and guide students to locate explicit information.	Literal questions are answered accurately.
Understands explicit information but fails to infer	Weak integration and inferential comprehension	Use causal questions, picture-text connections, and guided reasoning.	Simple text-based inferences are produced.
Progress does not match the diagnosed difficulty	Intervention is misaligned or insufficient	Review the profile and modify or intensify the strategy.	Improvement occurs in the originally diagnosed component.

6. Analytical Discussion

6.1 RDEM as an Evidence-to-Action Architecture

RDEM's principal analytical contribution is its reconstruction of the full evidence-to-action chain. Conventional early-reading assessment frequently collapses several distinct operations into one: a student performs a task, the teacher assigns a score or broad label, and instruction proceeds with little change. RDEM separates observation, measurement, diagnostic inference, profile construction, instructional decision, intervention, and outcome verification. This separation matters because each link can fail independently. A technically reliable score can still be instructionally useless when it does not identify the component causing difficulty, while a plausible diagnosis can still fail when it is not translated into a sufficiently specific teaching response. Research on curriculum-based measurement has long shown that repeated data produce stronger instructional effects only when teachers use explicit rules to alter instruction, rather than merely collect scores (L. S. Fuchs et al., 1989; L. S. Fuchs & Fuchs, 1986; Stecker et al., 2005)

The model therefore treats formative assessment as a change mechanism, not as a type of test.(Bennett, 2011; Black & Wiliam, 1998; Kingston & Nash, 2011))converge on the principle that evidence becomes formative only when it changes subsequent teaching or learning. RDEM operationalizes that principle by requiring every diagnostic conclusion to be paired with an instructional implication and a progress indicator. Feedback is consequently bidirectional: students receive information about the next attainable reading action, while teachers receive evidence about whether their instructional hypothesis was correct. Hattie & Timperley (2007) similarly distinguish feedback that merely reports performance from feedback that clarifies the goal, current position, and next step. In RDEM, feedback that does not alter the next decision is information, but it is not yet formative control.

6.2 The Analytical Role of Context

The responsive-context domain performs an interpretive rather than decorative function. It prevents identical scores from being treated as evidence of identical problems. Two students may read the same number of words correctly, yet one may be constrained by unstable grapheme phoneme mapping and the other by excessive caution, limited vocabulary, task unfamiliarity, or inconsistent exposure to print. Stake's responsive evaluation requires evaluators to examine antecedents, transactions, and stakeholder concerns before judging outcomes (Stake, 1975, 2004). This contextual layer is especially important because teacher judgments about reading difficulty often rely on partial indicators and informal impressions (Schmitterer & Brod, 2021). RDEM does not discard those observations; it subjects them to structured verification against component-level evidence.

Context, however, cannot be used as a substitute for diagnosis. Explanations such as limited home literacy, low confidence, or irregular attendance may be relevant, but they do not specify whether the student can discriminate phonemes, retrieve letter sounds, blend syllables, decode unfamiliar words, or understand a sentence. RDEM therefore imposes a dual-evidence requirement: contextual information explains the conditions under which



performance occurs, while diagnostic tasks test the hypothesized reading mechanism. The two sources must constrain each other. This prevents deficit interpretations that blame the learner or family, but it also prevents contextual narratives from becoming excuses for imprecise assessment. The resulting interpretation is stronger because it asks both what the student did and under what conditions that performance was produced.

6.3 Dynamic Profiles and Differentiated Instruction

Dynamic profiles are analytically superior to fixed labels because early-reading performance is multidimensional and developmentally unstable. The Simple View of Reading distinguishes word recognition from language comprehension, while later frameworks emphasize that word knowledge, lexical quality, and text integration interact during comprehension (Duke & Cartwright, 2021; Gough & Tunmer, 1986; Hoover & Tunmer, 2020; Perfetti & Stafura, 2014). A broad category such as ‘weak reader’ conceals these distinctions. RDEM instead represents the learner as a provisional pattern of strengths, errors, response conditions, and support needs. The profile must change when the evidence changes; otherwise, it becomes a disguised label.

This distinction also exposes a common weakness in differentiated instruction. Differentiation is often reduced to giving easier texts to lower-performing students and harder texts to higher-performing students. That practice varies task difficulty but may leave the underlying mechanism untouched. (Tomlinson, 2001, 2014) provides a framework for adapting content, process, grouping, and support, but the adaptation is only defensible when it is anchored in valid diagnostic evidence. For example, a student with weak phonological blending requires a different instructional sequence from a student who decodes accurately but lacks automaticity, even when both obtain the same oral-reading score. Oral-reading fluency is a valuable indicator, but it should not be treated as a complete diagnosis (Amendum et al., 2021; L. S. Fuchs et al., 2001) RDEM therefore uses flexible, component-based grouping and requires regrouping when progress data no longer support the original profile.

6.4 Decision Rules as the Operational Core

Decision rules are the operational core of RDEM because they expose the reasoning that normally remains implicit between assessment and teaching. A rule specifies the evidence pattern, the diagnostic interpretation, the initial intervention, the expected response, and the condition for changing course. This is more rigorous than matching a low score with a generic remedial activity. Response-to-intervention research demonstrates that screening, increasingly intensive support, and progress monitoring must be linked through explicit decision criteria (D. Fuchs & Fuchs, 2006) Similarly, formative-evaluation studies show that instructional adjustment, not measurement frequency by itself, accounts for the educational value of progress data (L. S. Fuchs & Fuchs, 1986; Stecker et al., 2005)

The rule must also preserve uncertainty. A repeated blending failure may suggest weak phonological integration, but the inference remains provisional until the student’s letter knowledge, task comprehension, and response to guided practice are examined. RDEM therefore treats intervention as a diagnostic probe: improvement after a well-aligned

intervention strengthens the working explanation, whereas nonresponse requires the teacher to question the construct, task, dosage, implementation, or contextual interpretation. This prevents a circular logic in which every failure is automatically attributed to the learner. It also prevents data abundance from being mistaken for evidence-based practice. Scores, checklists, and dashboards do not improve instruction unless they trigger a justified and observable change in teaching.

6.5 Responsive Outcomes and Model Falsifiability

RDEM adopts a stricter criterion for outcome interpretation than improvement in a composite score. An outcome is responsive only when it can be related to the initial need and to the instructional mechanism used. If the diagnosed weakness concerns decoding, the primary evidence of success should include more accurate or efficient decoding of relevant items, reduced prompting, and transfer to unfamiliar words not merely better performance on memorized material. (Ehri, 2005b; Gillon, 2018; Kilpatrick, 2015) all support the need to distinguish component skills and instructional targets in beginning reading. Consequently, the outcome stage asks whether the targeted component changed, whether the change generalized, and whether it increased independent reading.

This alignment principle makes RDEM falsifiable and self-correcting. When the expected change does not occur, at least five explanations must be considered: the initial diagnosis was incomplete, the task did not validly represent the construct, the intervention did not match the mechanism, the instructional dosage or fidelity was insufficient, or contextual barriers prevented engagement. The feedback loop therefore returns not only to instruction but, when necessary, to the original construct and evidence. This is a more demanding position than attributing nonresponse to low ability. It aligns with the broader science-of-reading argument that research must be translated cautiously into adaptive classroom practice rather than reduced to fixed prescriptions (Petscher et al., 2020; Seidenberg et al., 2020).

6.6 Conceptual Novelty, Limitations, and Validation Agenda

The conceptual novelty of RDEM lies in the integration of four functions that are typically treated as separate: Stake's context-sensitive evaluation, component-based reading diagnosis, formative evidence use, and adaptive instructional decision-making. None of these elements is new in isolation. The contribution is the proposed architecture that forces them to operate as one traceable cycle. Cognitive diagnostic assessment can produce fine-grained profiles (Mei & Chen, 2024; Min & He, 2022; Toprak & Çakır, 2021), but a profile alone does not determine instruction. Responsive evaluation can produce rich contextual understanding, but without operational indicators it may remain descriptive. Differentiated instruction can adapt teaching, but without valid diagnostic evidence it risks becoming intuitive grouping. RDEM's novelty is therefore relational and procedural rather than a claim to have invented its component theories.

The current study also has substantial limitations that must constrain its claims. RDEM was constructed conceptually from the classroom problem, theoretical synthesis, proposed instrument structure, and model logic; it has not yet demonstrated content validity, construct representation, scoring reliability, inter-rater agreement, decision-rule



accuracy, implementation feasibility, or causal effects on reading development. The current decision rules are reasoned hypotheses, not empirically calibrated thresholds. In addition, the model may impose cognitive and time demands on teachers, and its contextual interpretations could vary across evaluators. These are not minor technical issues: if profiles are unreliable or rules are applied inconsistently, the model could produce confident but incorrect instructional decisions.

A rigorous validation agenda should therefore proceed sequentially. First, experts in early reading, assessment, and evaluation should examine domain relevance, indicator clarity, theoretical coherence, and the alignment among evidence, interpretation, decision, and outcome. Second, diagnostic tasks and rubrics should undergo cognitive review and small-scale piloting to identify ambiguous instructions, floor or ceiling effects, and scoring disagreements. Third, classification consistency and decision accuracy should be tested against external measures and independent expert judgments. Fourth, classroom feasibility should be assessed through implementation time, teacher comprehension, fidelity, and acceptability. Only after those stages should a comparative or experimental study examine whether RDEM-supported decisions improve targeted reading components more effectively than existing practice. Until such evidence exists, RDEM should be presented as a testable model, not a proven solution.

Taken together, the sharpened analysis shows that RDEM should be judged by the integrity of its inferential chain. The model succeeds only when contextual evidence improves interpretation, diagnostic evidence differentiates mechanisms, profiles generate distinct instructional decisions, interventions are implemented with adequate fidelity, and outcomes confirm or revise the original hypothesis. A break at any point weakens the entire system. This criterion moves evaluation beyond the presence of instruments and toward the quality of reasoning that connects evidence with action.

CONCLUSION

The model's current status must nevertheless be stated precisely. RDEM is theoretically constructed, not empirically proven. Its value now lies in the coherence of its proposed evidence chain and in the testable questions it generates. Expert validation, instrument and rubric development, reliability analysis, feasibility testing, and classroom implementation are required before claims about effectiveness can be made. The next research task is therefore not to promote the model as finished, but to determine whether its diagnostic classifications and decision rules can be applied consistently and whether they produce more accurate instructional responses for beginning readers.

The model sharpens early reading evaluation in three ways. First, it distinguishes the source of difficulty across phonological awareness, decoding, fluency, and basic comprehension rather than relying on a broad reading label. Second, it introduces decision rules that connect diagnostic patterns with targeted instructional priorities. Third, it evaluates outcomes against the initial diagnosis, so that general score improvement cannot conceal a misaligned intervention. In this sense, RDEM positions evaluation as an instructional control system rather than as a terminal measurement activity.

The Responsive Diagnostic Evaluation Model was constructed in response to a structural weakness in early reading evaluation: assessment evidence is commonly generated without a reliable mechanism for translating it into instructional action. RDEM integrates responsive context, diagnostic antecedents, dynamic student-profile classification, differentiated instruction, formative feedback, and responsive outcomes in one adaptive cycle. Its core logic is that each evaluative stage must change the next instructional decision and remain open to correction by subsequent evidence.

ACKNOWLEDGMENT

The authors express their appreciation to the lecturers and members of Group 3 who contributed to the conceptual discussion and preparation of the original presentation on forms and strategies of assessment.

REFERENCES

- Amendum, S. J., Conradi, K., & Hiebert, E. H. (2021). Explaining reading variance by student subgroup: Should we move beyond oral reading fluency? *Journal of Research in Reading*, 44(4), 767–786. <https://doi.org/10.1111/1467-9817.12371>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the Simple View of Reading. *Reading Research Quarterly*, 56(S1), S25–S44. <https://doi.org/10.1002/rrq.411>
- Ehri, L. C. (2005a). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4
- Ehri, L. C. (2005b). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41(1), 93–99. <https://doi.org/10.1598/RRQ.41.1.4>
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3), 199–208. <https://doi.org/10.1177/001440298605300301>
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239–256. https://doi.org/10.1207/S1532799XSSR0503_3



- Fuchs, L. S., Fuchs, D., & Stecker, P. M. (1989). Effects of curriculum-based measurement on teachers' instructional planning. *Journal of Learning Disabilities*, 22(1), 51–59. <https://doi.org/10.1177/002221948902200110>
- Gillon, G. T. (2018). *Phonological awareness: From research to practice*. The Guilford Press.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10. <https://doi.org/10.1177/074193258600700104>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hoover, W. A., & Tunmer, W. E. (2020). *The cognitive foundations of reading and its acquisition: A framework with applications connecting teaching and learning*. Springer. <https://doi.org/10.1007/978-3-030-44195-1>
- Kilpatrick, D. A. (2015). *Essentials of assessing, preventing, and overcoming reading difficulties*. Wiley.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28–37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Mei, H., & Chen, H. (2024). Cognitive diagnosis in language assessment: A thematic review. *RELC Journal*, 55(1), 3–24. <https://doi.org/10.1177/00336882221122357>
- Min, S., & He, L. (2022). Developing individualized feedback using cognitive diagnostic assessment. *Language Testing*, 39(1), 90–116. <https://doi.org/10.1177/0265532221995475>
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22–37. <https://doi.org/10.1080/10888438.2013.827687>
- Peters, M. T., Hebecker, K., & Souvignier, E. (2022). Effects of providing teachers with tools for implementing assessment-based differentiated reading instruction in second grade. *Assessment for Effective Intervention*, 47(3), 157–169. <https://doi.org/10.1177/15345084211014926>
- Petscher, Y., Cabell, S. Q., Catts, H. W., Compton, D. L., Foorman, B. R., Hart, S. A., Lonigan, C. J., Phillips, B. M., Schatschneider, C., Steacy, L. M., Terry, N. P., & Wagner, R. K. (2020). How the science of reading informs 21st-century education. *Reading Research Quarterly*, 55(S1), S267–S282. <https://doi.org/10.1002/rrq.352>
- Schmitterer, A. M. A., & Brod, G. (2021). Which data do elementary school teachers use to determine reading difficulties in their students? *Journal of Learning Disabilities*, 54(5), 349–364. <https://doi.org/10.1177/0022219420981990>
- Seidenberg, M. S., Borkenhagen, M. C., & Kearns, D. M. (2020). Lost in translation? *Challenges in Connecting Reading Science and Educational Practice*. *Reading Research Quarterly*, 55(S1), S119–S130. <https://doi.org/10.1002/rrq.341>
- Stake, R. E. (1975). To evaluate an arts program. In *Evaluating the arts in education: A responsive approach* (pp. 13–31). Merrill.

- Stake, R. E. (2004). *Standards-based and responsive evaluation*. SAGE Publications.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences. *Reading Research Quarterly*, 21(4), 360–407. <https://doi.org/10.2307/747612>
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42(8), 795–819. <https://doi.org/10.1002/pits.20113>
- Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms (2nd ed.)*. ASCD.
- Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners (2nd ed.)*. ASCD.
- Toprak, T. E., & Çakır, A. (2021). Examining L2 reading comprehension ability: Developing a diagnostic test within the cognitive diagnostic assessment framework. *Language Testing*, 38(1), 106–131. <https://doi.org/10.1177/0265532220941470>
- Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.