Centering Teacher Voice: Using Social Validity to Support a Mandated Literacy Reform

Susan B. Neuman

New York University

Abstract

This study examines how social validity — teachers' perceptions of an intervention's feasibility, relevance, and value —shapes the successful implementation of a mandated literacy reform. In partnership with a large urban district, we supported first-grade teachers in integrating Geodes, a collection of 80% decodable, knowledge-building texts designed to bridge foundational skills and core English language arts (ELA) instruction. Using a design-based research approach we provided iterative coaching, lesson routines, and curriculum-aligned assessments to elevate the perceived value and usability of the materials. Findings showed that teachers' uptake of Geodes increased substantially with these supports, and classrooms that integrated the materials more consistently saw significantly greater gains in students' reading fluency. This work underscores that instructional materials alone are not enough: when educators are engaged as partners and supports are responsive to classroom realities, implementation strengthens and student outcomes improve.

Centering Teacher Voice: Using Social Validity to Support a Mandated Literacy Reform

Responding to stagnant literacy outcomes among more than half of its students, New York City in 2023 launched a sweeping initiative known as NYC Reads (n.d.). Marking a sharp break from the past when each principal could choose their school's reading program, the initiative mandated that all elementary schools adopt a literacy curriculum aligned with the "science of reading." As defined by the city, the science of reading refers to a multidisciplinary body of research that explains how children learn to read, emphasizing the importance of explicit, systematic instruction in phonemic awareness, phonics, fluency, vocabulary, and comprehension.

Over the next two years, each of the city's 32 community school districts were required to phase out the widely-used balanced literacy materials (e.g., (Calkins, 2015; Fountas & Pinnell, 2016) and implement one of three approved curricula—*HMH Into Reading* (HMH, 2023); *Wit & Wisdom* (Great Minds | Wit & Wisdom, n.d.) or EL education (EL, 2025). The new policy was in response to what then-Chancellor David Banks called a "flawed playbook" with "overlapping, contradictory, and sometimes flat-out bad guidance" for teaching reading (New Reading Curricula for City Schools, 2023).

Although the initiative was met with great enthusiasm from policymakers as well as cautious optimism from many educators, it also raised concerns about the loss of local autonomy and professional discretion (Collins & Vaughan, 2023). For more than two decades, New York City had followed a site-based model in which "every school is on its own" when it came to curriculum. The shift to a centralized, one-size-fits-all approach was unsettling for many

teachers, particularly those who had to abandon homegrown literacy projects or long-standing, well-loved teaching practices. Even with the extensive professional learning opportunities, and material supports accompanying the initiative, many veteran teachers worried that the mandate might override effective practices already in place.

These reservations underscore an important reality. Although the focus on evidence-based practice may be critical for improving outcomes, it is not enough on its own. The success of any reform will likely to depend on whether teachers find the approach feasible and worthwhile (Kozleski et al., 2021). In other words, even the most well-researched, and well-financed reform may falter if it lacks credibility among educators tasked with enacting it.

These dynamics point to a crucial construct in implementation science: *social validity*—the perceived fit, acceptability, practicality, and relevance of an intervention as judged by those expected to implement it (Wheeler & Carter, 2019; Wolf, 1978). Research consistently shows that even well-designed interventions struggle to take hold when they fail to align with stakeholders' values, priorities, or everyday realities. Without that alignment, implementation often lacks fidelity and long-term sustainability, regardless of positive evidence of efficacy (Clayback et al., 2023; Larson et al., 2020).

Recognizing its central role in enacting and ensuring the quality of an initiative, in this paper we explore how social validity influenced the adoption of a new mandated curriculum—from teachers' initial concerns about feasibility and fit, to the iterative supports we introduced in response, and how those evolving strategies contributed to shifts in teacher attitudes and, ultimately, to measurable gains in students' reading outcomes.

The Potential Role of Social Validity in Reducing the Pendulum Effect

The landscape of reading instruction has been said to resemble a pendulum, swinging between emphases on phonics and meaning-making, explicit and implicit instructional approaches, and structured programs versus teacher autonomy. These shifts are often driven by changing political priorities, evolving interpretations of research, and mounting public pressure for rapid educational improvement (Pearson, 2004; Hoffman et al., 2020). Yet despite their recurrence, such swings have seldom produced enduring improvements in students' reading outcomes. Evidence from the most recent National Assessment of Educational Progress (NAEP) ((NAEP, 2020)2025) only highlights this disturbing pattern, showing stark declines in both fourth and eighth grade reading scores, further exacerbating pre-existing downward trends.

Nevertheless, one could argue that the pendulum swing is due in part to educational reforms that have privileged externally derived standards of efficacy while neglecting the perspectives and situated expertise of classroom teachers. Top-down mandates often overlook the realities of instructional time, curricular alignment, and professional judgment, resulting in limited fidelity of implementation and diminished impact. These limitations are likely to fuel the next swing of the pendulum as policymakers look toward a newer 'solution' for the persistent reading challenges. This cyclical nature of reading reform, therefore, reflects a systemic problem: A history that has too often failed to recognize the complex social ecology of the classroom in which teachers' professional judgement and local knowledge are critical for success.

The concept of social validity offers a promising corrective. Broadly defined, it focuses on the degree of acceptance associated with a reform, listening to the voices of teachers tasked with enacting these changes. According to Wolf's classic definition (1978), it relies on three dimensions: (i) the social significance of the program's goals; (ii) the social appropriateness of

the program and (iii) the social importance of the actual outcomes. It takes into account the extent to which a treatment program is considered reasonable, justified and fair to implementers (e.g. teachers) as well as students. Studies (Biggs & Hacker, 2021; Rademaker et al., 2021) have linked implementation fidelity and quality of enactment to factors associated with teachers' positive attitudes to new programs. For example, Domitrovich and her colleagues (Domitrovich et al., 2008; Domitrovich et al., 2010) found that higher implementation dosage over time was related to teachers' initial beliefs and their perceptions of the new program's fit with their teaching style.

Importantly, however, this body of work also highlights a critical distinction: even when a program is considered effective, it may still be rejected by educators if it is not seen as appropriate or feasible within their instructional context (Wheeler & Carter, 2019). For a program to be sustained over time, it must not only demonstrate effectiveness but must also address problems that teachers find relevant and offer practices that are both manageable and personally meaningful in their daily work (Birken et al., 2020).

Consequently, although there is growing recognition of the importance of social validity, it has rarely been used to guide the adaptation of interventions prior to or during implementation (Larson et al., 2020). Instead, most studies assess social validity at the conclusion of an intervention (Kozleski, 2021). For example, in a review of special education literature, Snodgrass and colleagues (Snodgrass et al., 2018) found that nearly all reported evaluations of social validity were conducted post-intervention. Limiting social validity assessments to the end of a program, however, constrains its utility—it prevents timely adaptations and reduces opportunities to respond to teachers' concerns in real time. In this study, by contrast, we use

social validity from the outset to inform iterative adaptations, with the goal of aligning the intervention more closely with teachers' needs and classroom realities.

Moreover, within studies that include social validity, there has been great variation in the ways in which social validity is assessed. Some researchers have emphasized qualitative measures, including interviews, focus groups, and rating scales (Lyst et al., 2005; Leko, 2014). Others have argued for the need for more objective and psychometrically sound social validity assessments (Proctor et al., 2011; Rademaker et al., 2021). In our case, we chose to use multiple methods including surveys, focus groups, and informal conversations as well as feedback from progress-monitoring of children's reading fluency and comprehension to engage teachers in an ongoing conversation to support their effectiveness in enacting a mandated program.

Overview of the Mandated Program: The NYC Reads Initiative

As the largest school system in the United States, New York City Public Schools serves more than 900,000 students across nearly 1,600 public schools. Organized into 32 geographic community school districts, each district covers a set of neighborhoods and operates with a degree of local oversight within the broader citywide system.

Given the system's scale and diversity, the NYC Reads initiative for grades K–5 was launched in two phases: Phase 1 began in the 2023–24 school year across an initial set of districts while Phase 2 in 2024-2025 expanded implementation to include all the other districts the following year. District X (pseudonym), the focus of our research, was assigned to be in the second phase of the initiative. The District includes over 27,000 students and is ethnically diverse: 14.4% of the student population identifies as Asian, 12.5% African-American, 34.8% Hispanic, and 34% white; 16% identify as English learners, and 23.5%, special needs. Over

three-quarters of the students are economically disadvantaged, and qualify for free and reduced lunch (FRL).

Like those districts in the first phase of implementation, District X was required to adopt a number of instructional shifts as part of the NYC Reads initiative (NYC Literacy Shifts 2023-2024). Designed to advance system-wide improvements in early literacy instruction by aligning classroom practices with the science of reading, schools throughout the districts were instructed to move away from 'less researched practices' to 'research-informed practices.' These research-informed practices now required teachers to provide systematic, explicit instruction in phonics, fluency, comprehension and writing, assess students' progress through universal screening, differentiate instruction, and adopt a research-based knowledge-building reading program, selected from the small menu of mandated programs for K-5 classrooms. As part of this effort, each district was to choose a core program (among the three mandated choices), and an outside professional learning organization who would provide training and implementation supports to teachers.

District X selected Wit & Wisdom as its core reading program. Known as a knowledge-building curriculum, the program emphasizes reading comprehension, vocabulary, writing, and speaking through thematic units that include narrative and information texts. Unlike the other two mandated options for the city (e.g. Into Reading; Expeditionary Learning), however, Wit & Wisdom does not include an explicit foundational skills component (phonemic awareness and phonics), but is designed to be complementary with a separate, systematic phonics program. In District X, the curriculum is paired with Fundations (Wilson Language Training) most frequently, a systematic and explicit phonics program. Both are required to be implemented in K-3 classrooms.

But interestingly, an additional program designed to serve as an opportunity for students to practice their foundational skills, and develop background knowledge essential for the core reading program was not part of the initial mandated materials. *Geodes*, developed by Great Minds (n.d.), is a collection of knowledge-building texts designed to provide students with meaningful reading experiences that are both content-rich and 80% decodable, fully aligned with foundational skills instruction and core ELA curricula. Though not a part of the initial mandate, Geodes was identified by the district as a strategic complement to ensure continuity and application across the literacy block. Recognizing that students needed opportunities to engage in daily connected text reading aligned with their phonics instruction, they found that Geodes could serve as a critical bridge between the foundational skills program (*Fundations*) and the core knowledge-building curriculum (*Wit & Wisdom*). In response, District X chose to recommend that Geodes become embedded into Tier 1 instruction, ensuring that all students would have access to structured, decodable reading experiences within the literacy block.

To support this shift, the district sought a collaborative partner to assist in promoting effective classroom use of *Geodes*. While schools were already receiving job-embedded coaching from an external provider for *Fundations* and *Wit & Wisdom*, they had received no structured guidance for integrating Geode**s** into daily instruction. Our university-based research team was invited to serve as an independent partner in this effort. In collaboration with district leaders, our goal was to support the instructional enactment of Geodes as part of Tier 1 literacy instruction to ensure that all students received high quality differentiated instruction. As part of this partnership, we also agreed to measure children's progress throughout the year, and to provide ongoing feedback to the schools, and district office.

Research Design

We conducted this research as a design experiment, with a focus that was both pragmatic and theoretical. Design experiments work in real-world educational settings and involve iterative cycles of design, implementation, observation, and refinement, aimed at both improving practice and generating theory about learning (The Design-Based Research Collective, 2003). Our focus was to understand the learning ecology of classrooms, particularly the kinds of supports teachers would need to adopt a new reading program that stood in stark contrast to their previous instructional practices. Design-based research functions as a "test-bed" for innovation, beginning with initial conjectures about how learning and implementation might unfold (Cobb et al., 2003). In our case, in collaboration with the district, we were to offer coaching support specifically on Geodes in the form of observation, modeling, and demonstration lessons. However, as the study progressed, we developed more targeted conjectures, leading to iterative cycles of design, implementation, and refinement.

We drew on complementary strands of implementation and improvement science to frame the design and analysis of our work (Domitrovich et al., 2010). We were particularly influenced by Fixsen and colleagues' Active Implementation Frameworks (Fixsen et al., 2005), which emphasize the importance of core implementation drivers—competency, organizational supports, and leadership—in supporting the effective use and sustainability of innovations. These concepts helped us attend to the systemic and organizational conditions shaping teachers' uptake of the new reading program. At the same time, we were guided by principles of improvement science as articulated by Bryk and colleagues (Bryk et al., 2015), which foreground the importance of iterative testing, practitioner learning, and adaptation to local contexts. In line with this tradition, we viewed implementation not as a linear process of program delivery, but as an

ongoing, collaborative process of refining supports and responding to teachers' evolving needs. Together, these frameworks provided a lens for understanding how both system-level drivers and classroom-level adaptations shaped the enactment of the curriculum in practice.

A distinctive feature of our approach was the integration of social validity not as a post-hoc evaluation but as a guiding framework throughout the process. By systematically attending to teachers' perceptions of the intervention's relevance, feasibility, and usefulness, we were able to adapt both the content and delivery of the program in real time. In doing so, we positioned teachers not merely as implementers but as co-designers whose insights shaped the evolution of the program. This use of social validity strengthened the responsiveness of the design process and contributed to its potential for rapid, practice-centered impact—addressing the kinds of challenges educators face in their day-to-day work. By situating our research in authentic classroom contexts and making teacher input central, we sought not only to improve implementation but to enhance the likelihood that the intervention would be seen as meaningful, manageable, and worth sustaining.

Building on this prior work, we designed the current study to examine how teachers' perceptions of social validity evolved over time as they implemented a newly recommended addition (e.g. Geodes) to their mandated literacy programs, and how these perceptions shaped classroom practice and student learning. Specifically, we addressed the following research questions:

1. How did first-grade teachers perceive the feasibility, relevance, and value (social validity) of Geodes at the start of implementation?
2. How did iterative, design-based supports shape teachers' perceptions of social validity during the year?

3. To what extent did shifts in perceived social validity influence teachers' level of Geodes use and integration within Tier 1 instruction?
4. How did different levels of Geodes implementation relate to changes in students' reading fluency progress over the school year?

To address these questions, we employed a mixed-methods design, combining teacher surveys, interviews, classroom implementation data, and student progress measures collected over the course of the school year.

Methods

Sites and Participants

Fourteen of the district's 18 elementary schools participated in this study; four dual-language schools had received permission from the Department of Education to adopt a different program and were therefore not included. Within the 14 participating schools, our study focused on first-grade classrooms**,** a critical, high-stakes year for early literacy outcomes. Given that it was the first year of NYC Reads implementation for District X, teachers were grappling with the introduction of two mandated programs in phonics and knowledge-building, along with a proposed additional recommended program—Geodes, for building reading practice, fluency and comprehension. Understanding the scope of these instructional changes, we anticipated that teachers' perceptions of feasibility**,** fit**,** and effectiveness would be particularly salient in first grade, making it an informative context for examining social validity in relation to mandated literacy reform. In total, 57 teachers from 35 first-grade classrooms participated in the study, including 15 integrated co-teaching (ICT) classrooms and 16 general education classrooms (see Table 1).

Insert Table 1 about here

First Steps in Implementation

We introduced the Geodes program to first-grade teachers in late summer 2024 through a two-hour virtual workshop. During the session, we reviewed the purpose of the materials and how they are designed to give students an opportunity to practice skills acquired through both their foundational skills program and their knowledge-building curriculum. The Geodes materials are described as "readables," developed to align to 80% of students' current level of decoding instruction (e.g. Fundations to Geodes, lesson-to-text-match, (Hiebert, 2024). Within the first grade series, there are 64 unique titles, organized in four modules, each with four text sets. In contrast to highly controlled decodable texts, in which as many as 90% of words reflect explicitly taught phonetic patterns, the Geodes texts incorporate both phonics-based elements and vocabulary drawn from the Wit & Wisdom curriculum materials. This structure allows for the inclusion of more varied linguistic features, and informational and narrative elements than is typically found in conventional decodable texts. We also described ways that teachers might differentiate instruction for children with varying skills based on their end-of-year screening scores, and the possible frequency and duration of these initial lessons.

Initial Implementation: First Iteration

Our initial workshop had limited immediate impact. Follow-up visits to classrooms revealed that Geodes materials often remained unopened—stored in boxes, closets, or placed alongside an array of both newly introduced and longstanding materials. In conversations with teachers, a consistent theme emerged: many felt overwhelmed. Because the other two programs were mandated while Geodes was only recommended, teachers necessarily prioritized their instructional planning around the required programs.

At the same time, both district leaders and teachers recognized the importance of providing students with opportunities to read connected text. In response, and in collaboration with District X's instructional team, we developed an initial plan for job-embedded coaching to support the integration of the new materials and to promote greater instructional coherence across the literacy block. We employed four coaches, each of whom held a master's degree (or, in one case, a doctoral degree) and had five or more years of experience teaching in elementary grades. The plan included weekly classroom visits, modeling, and collaborative planning sessions, with particular emphasis on helping teachers navigate the intersections among Fundations, Geodes, and Wit & Wisdom and make informed instructional decisions across the three programs. Based on evidence from previous research on the efficacy of coaching (Kraft et al., 2018), our plan was to provide a high dosage of 30-minute sessions (12-weekly) per classroom, and focus primarily on demonstrating practices for teachers within their actual instructional time (Schachter et al., 2025).

To prepare for these sessions, we constructed a brief online 7-item survey to better understand teachers' needs as they began implementation. Items included questions about their familiarity with the materials, schedule, plans for differentiating instruction, and potential barriers for implementation. We also touched base with teachers at each school to finalize our next steps.

Findings from our First Iteration: Teacher Perspectives on Barriers to Implementation

We imported the survey and field notes from our coaching visits into MAXQDA, the qualitative software program for analysis. Using a code of 'challenges to initial implementation,' we conducted an iterative thematic analysis (Braun & Clarke, 2006), generating a set of emergent categories reflecting the primary barriers to teacher uptake of Geodes. The coding

process involved multiple cycles of review, with codes identified and then refined by two research assistants working independently who then compared their analysis of segments across schools and grade levels. To assess the reliability of our coding scheme, two researchers independently coded 20% of the transcripts. Intercoder reliability was calculated using Cohen's Kappa, yielding an agreement coefficient of $\kappa = .82$, indicating strong consistency (Landis & Koch, 1977). Discrepancies were discussed and resolved through consensus before coding the full dataset**.** The resulting themes were then synthesized with findings from our survey to construct an integrated picture of the first iteration of implementation.

Both the survey and the coaching field notes were highly consistent, revealing a set of challenges that shaped teachers' early engagement with the Geodes materials. Across both data sources, several recurring themes emerged: concerns about time constraints**,** uncertainty about instructional integration**,** logistical barriers**,** and initial skepticism**.** Below, Table 2 synthesizes these challenges that teachers faced during this first iteration of implementation.

Insert Table 2 about here

While there was interest and goodwill among many of the educators, they were faced with a complex set of barriers, both structural and perceptual that hindered the use of these materials. One major theme, as shown in Table 2, reflected a sense of competing demands. Many teachers felt stretched thin by the sheer number of new initiatives and expectations. "These new materials are taking so much time. I don't have the time to implement these lessons and I also don't know how they fit into my day." Compounding these challenges were logistical barriers. In a number of schools, Geodes materials were still packed in boxes or sitting on their shelves virtually unused. As one instructional leader described it:

> "Overall, these are a seasoned group of teachers who are stressed with implementing the mandated program. Most have had Geodes in their classrooms for the last year, but have not implemented them. The other two programs take up three periods, and with two periods dedicated to lunch and recess, and math, there's not much time for small group literacy work."

Relatedly teachers voiced uncertainty about how to use the new materials effectively, especially in relation to the other mandated programs. For example, as one teacher put it, "The Geodes are based on alignment with the other two programs, yet they don't allow for variations across the skill levels of students." Added another, "Geodes are simply too difficult for most of my students." And another, "I think Geodes connections to the other two programs are just superficial. I don't see a connection." While some educators expressed skepticism about the alignment, these perceptions often reflected a lack of structured support and clarity rather than issues with the materials themselves. Therefore, without guidance on how Geodes could align with their current mandated literacy block, many felt uncertain about how to proceed.

But perhaps most telling from a social validity perspective was a sense among teachers of loss and frustration around the dismantling of prior instructional practices. There was clearly some initial skepticism and resistance, rooted not in hostility to the program itself, but the purpose for these massive changes in instructional routines. In particular, many had invested years in the Fountas & Pinnell leveling system and the Teachers College Reading and Writing Project (TC) approach. The district's new direction required them to retire extensive collections of leveled books, a resource they had carefully built and used to guide instruction. Teachers also worried about how to communicate children's reading progress to parents without using familiar level-based language (e.g., "your child is reading at Level J"), which had become embedded in

their routines and parent conversations. As one teacher put it, “I feel very frustrated that our system for assessing children and sending home books has been taken away from us with nothing put in its place.” These disruptions compounded feelings of uncertainty and loss of professional agency, underscoring the importance of attending to the perceived *fit*, *practicality*, and *relevance* of new practices, not only in abstract terms but in relation to the everyday tools and language teachers relied upon.

Collectively, these concerns highlighted several barriers to implementation. Teachers faced practical challenges, such as limited instructional time and uncertainty about how to align the new materials with other mandated programs. They also grappled with the loss of familiar practices which disrupted established routines and professional identities. As one instructional leader told us:

> “My teachers are mostly struggling with how to move from guided and leveled reading lessons. They want help on learning how to introduce a new book. They want help in understanding how to group students. They want to know when to do a whole group lesson and when to do a small group lesson. They want a structure.”

Finally, teachers sought greater clarity on how to structure instruction with Geodes while ensuring that all children received equitable Tier 1 access to the texts’ content. From our perspective, therefore, addressing these barriers called for supports that transcended conventional coaching models, offering a collaborative “helping hand” to assist teachers in navigating the complexities of this instructional shift.

Next Steps in Implementation

Based on this evidence, we adopted a revised plan for working with teachers, one that maintained a central focus on improving student learning. This perspective was informed by Hasbrouck and Michel's student-focused coaching model (Hasbrouck & Michel, 2022), which positions coaching as a means to promote student success. Recognizing that not all teachers were inclined to participate in traditional coaching, we reframed our approach as a collaborative effort to support students' reading growth, a shift that we hoped might foster greater receptivity and engagement. We emphasized three key areas in our revised approach: (i) centering student learning through assessment and feedback; (ii) promoting instructional coherence and usability; (iii) strengthening teacher engagement and agency, which we describe below.

Centering Student Learning through Assessment and Feedback

A central feature of the revised coaching model was the emphasis on student learning, beginning with tools and strategies to monitor children's reading progress. Although the District required teachers to administer universal screeners such as Acadience (Good et al., 2011) three times per year, these assessments offered limited insight into how students were actually progressing within the curriculum materials themselves, and how this could help to differentiate instruction.

To address this gap, we developed a curriculum-aligned assessment protocol that was approved by the University's and NYC's Institutional Review Boards. Parents of all participating students provided informed consent prior to data collection. We then randomly assessed approximately 10 students per class for a total of 407 first graders.

To generate a more instructionally sensitive measure of students' reading progress with the Geodes materials, we designed a curriculum-aligned assessment protocol. To avoid adding to teachers' workload, a research team of four masters' level students in applied psychology conducted all individual student assessments. Drawing on the expected pacing of the Geodes sequence, we selected an unfamiliar text that had not yet been introduced in class. Each child was asked to read the passage aloud for one minute. If the child did not complete the passage within that time, the remainder was read aloud by the assessor to ensure consistent comprehension opportunities. Immediately afterward, the student was prompted to retell the story, and the retelling was audio recorded and analyzed. This protocol yielded multiple data points: (1) words correct per minute (WCPM), benchmarked against Hasbrouck and Tindal's national oral reading fluency norms (Hasbrouck & Tindal, 2017) to gauge decoding accuracy and rate; (2) retelling quality, assessed using a story grammar framework to capture comprehension of narrative structure (e.g. setting, characters, plot, events, resolution) (Mandler & Johnson, 1977); and (3) lexical density (Ratner et al., 2024), calculated based on the proportion of content-specific words used in the retelling. A summary was provided to each teacher, highlighting what level of supports various groups of students should receive. Collectively, these metrics provided a multidimensional view of students' decoding, fluency, and comprehension in relation to the instructional texts, and could be used by teachers to help in differentiating instruction.

Promoting Instructional Coherence and Usability

Next, we sought to address teachers' concerns about the questions of Geodes' fit in the classroom, and issues tied to the instructional integration and coherence of the programs. In response, we first created an alignment chart, indicating the skills to be taught in Fundations, its

application to the Geodes instructional materials, and the instructional targets in Wit & Wisdom to serve as a visible guide of the alignment between programs.

In doing so, we also recognized that it was important to highlight the specific instructional purpose of the Geodes materials, how it bridged skills taught in the other programs, and provided an opportunity for students to put their emerging decoding skills to practice in connected, engaging texts as a means to develop fluency and comprehension. However, it was clear that some teachers felt that the words in the Geodes books were too advanced for some children in the classroom, and appropriate for only their highest group of readers. Therefore, to ensure that *all* children had access to these materials as part of their Tier 1 core instruction (e.g. comprehensive reading instruction), we developed a structured approach that combined whole-group and differentiated small-group teaching. The instructional routine included introducing the Geodes book to the whole group at the beginning of the week to highlight the key vocabulary, model fluent reading, and engage all children in listening and responding to comprehension questions. We created power-point slides of the stories for children to see the print as the teacher read along, integrating questions from the teachers' guide to facilitate discussions. We then developed small group lesson plans, describing high-leveraged routines for children at three skill levels including finger-point reading, phrasing, repeated reading, and close reading of text. These activities were designed to address the purposes of the materials and to help alleviate some of the concerns about the program's applicability for students with different instructional needs.

Strengthening Teacher Engagement and Agency

Virtually all teachers, through our email exchanges with them or visits to the schools, seemed to appreciate these new resources. Nevertheless, adopting a new instructional model proved difficult for many teachers, particularly those who had long valued the familiarity and autonomy afforded by guided reading. It was not resistance as much their lack of comfort and confidence in comparison to previous practices.

Yet it led to a number of concerns. Despite the district's shift, we found that some teachers were blending programs, leading to a 'little bit of this, and a little bit of that.' The coherence of instruction in their classrooms fell flat. As one teacher put it, "there are benefits to both approaches," while another referenced "choosing elements from both"—comments that pointed to an ongoing negotiation between old routines and new mandates. These compromises not only undermined instructional coherence but also hindered full adoption of the new materials as intended for Tier 1 use. Compounding this, several teachers misunderstood the purpose and potential of Geodes, characterizing them as texts "for the lower students." This misconception, unfortunately, limited use in some classrooms, despite the fact that the materials were designed to offer equitable access to all students through Tier 1 instruction. Together, these issues highlighted how instructional challenges, whether due to misperceptions, misalignment, or student response, could weaken both teacher buy-in and the effectiveness of implementation.

Recognizing these challenges, we sought to address both the instructional gaps and teachers' lingering discomfort with the new materials by shifting our approach, moving beyond traditional observation-and feedback coaching to working directly with students in classrooms. In some classrooms, our support consisted of just a few targeted coaching sessions in which we demonstrated key routines and teachers quickly felt ready to implement them

independently. In other settings, we worked more intensively, engaging in side-by-side small-group instruction over multiple visits until teachers felt confident enough to try the new approaches on their own. As one teacher reflected, “having someone right there to show me how it could work with my students made it feel possible — not just another thing to figure out.” By working directly with small groups, we were able to make visible the kinds of reading behaviors and vocabulary challenges that students faced, and teachers could see how these insights connected back to what they were trying to accomplish in small-group lessons.

We also provided some step-by-step descriptions of common practices such as repeated reading, echo and choral reading, and phrasing, de-composing these practices to become high leverage routines. We introduced short, focused small-group lessons; for example, 12-minute fluency-focused sessions to demonstrate how Geodes could be used to supplement phonics instruction and build reading fluency without adding undue burden to teachers’ schedules. When appropriate, we modified lesson materials to ensure accessibility for all students, by providing simplified versions of texts or photocopied pages that students could take home to practice with their families. In parallel, we emphasized the use of progress-monitoring data to guide instructional decisions, encouraging teachers to use flexible grouping and differentiated support based on students’ emerging needs.

Importantly, this iteration emphasized collaboration and flexibility rather than rigid prescription. Throughout the year, we averaged 7 in-person coaching sessions per classroom, with the number of sessions ranging from 2 to 16 sessions depending on each school’s needs; this was supplemented by weekly online updates. By co-constructing models of use and creating space for teachers to adapt lessons to their own context, we aimed to address the barriers related to teacher comfort, perceptions of fit, and instructional coherence by strengthening the perceived

relevance, value and feasibility of integrating Geodes more fully into Tier 1 instruction. (Fixsen et al., 2005; Kraft, Blazar, & Hogan, 2018).

Analyzing Social Validity and Implementation Outcomes

To understand the influence of our second iteration and the additional supports introduced to strengthen social validity, we drew on multiple data sources collected toward the end of the year. First**,** we administered a 7-item post-program survey to examine the average frequency of Geodes use, grouping arrangements, and the approximate amount of instructional time devoted to whole-group and small-group lessons each week.

We reviewed these results in consultation with our coaches to triangulate these data. Based on this information, coaching reports, and classroom visits, we developed a three-level typology to reflect the multiple factors that accounted for use, ranging on a scale from consistent integrated use (Level 1) to minimal or no use (Level 3). For example, a number of teachers demonstrated consistent use, embedding Geodes across whole class, small group, and independent settings on a near-daily basis, with strong alignment to other core materials. A larger proportion fell into the limited use category, using Geodes regularly in small groups 2–3 times per week, often to reinforce foundational skills or build fluency. Finally, a notable group fell into the minimal/no use category, where Geodes were either not used for instruction or made available only informally in classroom libraries. Teachers in this category cited barriers such as time constraints, perceived difficulty of texts, or a lack of alignment with other mandated programs. Each classroom was then classified according to this typology to indicate its level of Geodes implementation.

Second, we conducted 40-minute focus groups with our focal teachers in each school to better understand their perceptions of the materials' relevance, value, and feasibility. Discussion prompts explored how well the materials met students' needs, teachers' interest in continued use, and whether they preferred the new approach over previous instructional practices. We also sought to understand whether or to what extent the additional supports we provided affected the feasibility, and usability of the materials for their students. Transcripts of these interviews were imported and analyzed using our MAXQDA software. Open-ended responses were coded thematically to identify shifts in attitudes and to highlight factors that contributed to greater perceived fit or persistent barriers.

Third, we analyzed student progress monitoring data to explore patterns in engagement and learning outcomes. Once each classroom was classified according to its level of implementation, we compared these classifications with students' collective progress monitoring results on pre- and posttest progress monitoring gains to examine whether differences in dosage were associated with differences in student outcomes. While not aiming to establish causal relationships, this descriptive analysis could provide an important context for teachers' reports and help to illustrate how more consistent implementation aligned with improvements in foundational skills and content knowledge. Together, this multi-level analytic approach was designed to address our research questions about the implementation of a mandated program and its perceived fit within Tier 1 instruction.

Results

In the following section, we present our findings, organized to address our core research questions about the feasibility and perceived social validity of implementing Geodes as part of Tier 1 instruction. Our goal was not only to determine whether teachers used the materials, but to

understand how their evolving perceptions of *relevance*, *value*, and *feasibility* shaped use, key elements of social validity. For each question, we integrated illustrative quotes through our observations and focus groups to highlight teachers' perspectives on what helped, what constrained their use, and how they envisioned sustaining or improving implementation moving forward.

Changes in the Use of Materials

Our initial survey indicated that approximately 77% of the teachers did not use the materials in teacher-led instruction. Several teachers indicated that they placed these books in literacy centers, or made them available during transition time. As one teacher put it, "Geodes are used when I can get to them—maybe once a week?" Other teachers found the books too difficult for their students, while still others claimed that they were "not challenging enough for my above grade level readers." More often than not, as described above, teachers listed the barriers of lack of time, and appropriateness for their students to account for their limited use.

Insert Table 3 about here

As noted in Table 3, teachers' use of materials changed substantially by the end of the year; 83% of classrooms were using the materials to some degree, with 46% reporting consistent use. However, 42% of classrooms still fell into the category of limited or occasional use, suggesting that while teachers were increasingly recognizing the value of these materials, many continue to use them only sporadically. About 17% of classrooms reported minimal or no use, echoing earlier concerns about barriers such as lack of time and uncertainty about the texts' appropriateness for their students. Together, these results suggest that for many, the materials remained a supplemental resource rather than a fully integrated part of Tier 1 instruction.

Most Successful Supports

Nevertheless, this increase in use suggests that many teachers were willing to integrate these materials, at least to some extent, into their instructional program. To build on this momentum, we examined which forms of support were most effective not only for implementing Geodes use but also for informing how mandated programs like this can be implemented with greater social validity. Three key factors emerged from the data that appeared to strength teachers' professional learning and practical integration of materials. These include:

Clear, ready-to-use lesson plans and slides. Teachers' manuals are often overloaded with ideas and activities, making it challenging for teachers to discern which elements are essential for effective instruction and which can be set aside. Many teachers noted that having a clear, streamlined lesson plan served as a practical roadmap, reducing planning time and making implementation feel far more manageable. As one teacher explained, *"The structure of the first sample model lesson you gave us made it much less intimidating — it wasn't this huge lesson plan. It was so simple: access prior knowledge, review vocabulary, here's the book for the read aloud.* In short, simplifying the lesson design and providing clear, usable templates and slides helped to reduce planning barriers and increased the feasibility of integrating new materials into daily instruction.

Hands-on student-focused coaching. Another factor that supported implementation was the hands-on, student-focused nature of the coaching. Instead of concentrating on teacher performance, coaches often worked directly with small groups of students, modeling lessons in real classroom contexts. Sometimes this simply meant *"having an extra adult who's working with one group,"* which teachers described as invaluable for differentiating instruction and keeping students engaged. But it was often more than simply having another adult in the room.

Teachers seemed to appreciate a collegial non-evaluative voice, someone who could share thoughtful observations and reflections about students' reading behaviors, insights that they might have missed on their own. One teacher explained, *"I found, after you worked with a group and shared what you noticed about the kids, that was really helpful — those are things we weren't seeing."* By keeping the focus on students' learning, this coaching approach helped teachers feel supported rather than evaluated, which appeared to strengthen the perceived relevance and feasibility of the materials — important conditions for teachers' willingness to continue integrating them into practice.

Progress Monitoring and Feedback. Another factor that appeared to support teachers' use of the materials was the regular progress monitoring and targeted feedback that made students' growth visible and instructionally relevant. Progress monitoring was directly tied to the curriculum materials themselves, helping teachers see how students were progressing with specific reading fluency and comprehension skills in context. After each assessment, teachers received a concise "report card" showing student progress across the year — from beginning to middle to end. Several teachers noted that this made student growth feel more concrete and immediately applicable to their planning. As one teacher described, seeing students' gains and knowing which skills still needed attention helped them adjust instruction more purposefully. This cycle of actionable feedback seemed to increase the perceived value of Geodes as a tool for informing instruction, and the visibility of students' progress was motivating for teachers, strengthening their sense that the effort to integrate the materials was worthwhile.

High-leverage routines. A final factor that appeared to support teachers' implementation was the development of high-leverage routines that targeted students' diverse needs while streamlining planning. The team designed specific lesson structures — such as *finger-point*

reading for students struggling with decoding, *phrase reading* for those who read haltingly, and *close reading routines* for students needing extra comprehension support. These step-by-step guidelines drew on research-based practices like repeated reading and gave students consistent ways to engage with the texts. Teachers found this structured approach especially effective, noting that it made it easier to use Geodes meaningfully across different instructional formats. As one teacher put it, *"We saw the value in the books from the lessons you did. We liked the idea of having a book a week and using the structure you had with the whole and small groups."* Another noted simply, *"Teachers found the structured approach of using Geodes effective, especially when used in whole-group settings."* By embedding clear routines into both whole-group and small-group work, teachers could differentiate more effectively while cutting down on planning time, an important benefit given their competing demands.

Taken together, these factors suggest that when supports are practical, student-focused, and aligned with teachers' real instructional needs, they helped to increase the perceived relevance and feasibility of a mandated program. Although we make no claims for causality, we can cautiously speculate that they may have contributed to their greater implementation within the realities of classroom practice.

How Material Use Relates to Student Outcomes

In our final analysis, we examine the relationship between material use and student outcomes in reading fluency. As shown in the Figure, our results showed a strong relationship between how consistently teachers used Geodes and student growth in reading fluency. Across all classrooms, students made measurable gains from fall to spring, but the size of those gains differed meaningfully by level of use. Classrooms with the highest levels of Geodes integration

(Level 1) showed the greatest average improvement in reading fluency, while students in classrooms with minimal or no use (Level 3) made smaller gains on average.

See Figure 1 here

To examine whether students' reading fluency gains differed by the level of Geodes integration, we conducted an analysis of covariance (ANCOVA) controlling for pretest scores on on posttest Words Correct Per Minute (WCPM) scores. The analysis revealed a significant effect of usage level on student gains, F (2, N) = 4.33, $p < .013$, with an effect size of $\eta^2 = 0.068$, indicating that approximately 7% of the variance in student gains was explained by differences in Geodes use. Post-hoc comparisons using Tukey's HSD test showed that students in classrooms with consistent Geodes use (Level 1) made significantly larger gains than those in classrooms with limited (Level 2) or minimal/no use (Level 3). Students in classrooms with limited use (Level 2) also outperformed those with minimal use (Level 3).

Overall, this pattern suggests that more consistent integration of Geodes was associated with stronger student growth in reading fluency. Further, it provides promising evidence that the greater integration of materials and instructional dosage in reading practice could be linked to measurable improvements in students' reading development.

Discussion

Social validity captures how relevant, acceptable, and important a program feels to those participants directly affected by it (Wolf, 1978). Increasingly recognized in implementation science (Birken et al., 2020), it reflects not only whether an intervention works, but also how well it fits within the lived experiences and values of educators and students. Studies suggest that practitioners are more likely to implement programs with fidelity — and to sustain new practices

over time — when they find them meaningful, feasible, and aligned with both their students' needs and their own professional values (Wheeler & Carter, 2019).

Therefore, the purpose of this study was to understand what it takes to implement a mandated literacy initiative in ways that teachers might perceive as feasible and worthwhile to benefit student achievement. The research was conceptualized as a design-based process, characterized by iterative refinements as we attempted to respond to teachers' needs and align the program with classroom realities.

We began our work by offering job-embedded coaching to teachers, drawing on research-based principles that emphasized classroom observations, timely feedback, and demonstration lessons — approaches widely supported in the professional development literature (Joyce & Showers, 1996; Kraft, Blazar, & Hogan, 2018). However, recognizing the barriers posed by limited time and the need to maintain teacher buy-in, we adapted our model in the second iteration to incorporate student-focused coaching (Hasbrouck & Michel, 2022). Among other activities, this shift had coaches working directly with students in the classroom, helping teachers to differentiate instruction in real time, trying out high-leverage strategies with students of diverse needs.

These practical supports, which included lesson plans, slide decks, routine templates of high leverage practices, assessments and opportunities for data-driven reflections, as well as working directly with children in small groups seemed to help break down barriers that might otherwise limit implementation. These teachers were not resistant to change; rather, they faced the very real challenge of having too many requirements placed on them in too short a time. This non-evaluative 'helping hand' provided by coaches served as a temporary scaffold, a time for teachers to catch their breath in the face of enormous changes in practice. As one teacher put it,

"We get things thrown at us, and we have to make it work." The student-focused coaching was often short-term for teachers who began to gain confidence in using the materials in small groups; for others, it extended over a longer duration, with up to twelve sessions. This flexible, responsive support helped many teachers begin to integrate these lessons into their literacy block and sustain the practices independently over the year. By adjusting the intensity of coaching to fit teachers' needs, we aimed to strengthen the perceived acceptability and relevance of the program for student learning. Still, however, it is worth noting that these efforts were not entirely successful. A fair number of teachers; 22% of teachers reported only minimal to no use of materials. It highlights the limits of supports and what is perceived as feasible and worthwhile among teachers who serve students with different needs.

Importantly, this study shows that even research-based materials such as the Geodes benefit from implementation supports that prioritize feasibility and usability. While the texts themselves are intentionally designed to align with foundational and knowledge-building programs, their potential was most fully realized when accompanied by professional learning supports. These supports served not only to increase uptake, but to drive measurable gains in students' reading fluency, suggesting that implementation quality is a key lever in maximizing the effectiveness of these reading resources.

It raises questions about professional learning and coaching in specific. Professional learning opportunities that include specialized training, online webinars, and workshops may introduce teachers to the research underlying the science of reading, and inspire them to try new ideas based on this acquired knowledge. These trainings are often supplemented with job-embedded coaching with experts in the field who model research-based practices and work with teachers to incorporate them into their own classrooms. In contrast, our approach positioned

coaches not as distant experts but as collaborative partners with a target-like focus on student learning. Therefore, rather than simply modeling practices for replication, we gave them tools, then checked to see how these supports were working in practice. By adapting lessons in real time and responding to students' diverse needs, coaches helped teachers align new routines within their existing context. In this respect, both coaches and teachers acted as co-learners, jointly engaged in efforts to maximize students' reading progress. By centering on students' needs, these results may add to the corpus of research on coaching and may help to clarify aspects of coaching practice that affect students' progress in reading frequently under-specified in previous meta-analytic reviews (Schachter et al., 2025).

Our findings also speak to the implementation of mandated curriculum programs, a policy increasingly adopted by states and districts across the U.S (e.g. 25 states to date (Neuman et al, 2023). Based on our research, it is clear that a one-size-fits-all approach will not be successful. Recognizing the wide range of student skills within classrooms and the critical role of Tier 1 core reading instruction, programs will need to be adapted to address students' diverse needs, particularly for those who require more differentiated support. Teachers, therefore, will need guidance in making structured adaptations that maintain the integrity, intent, and focus of the program (Kim et al., 2017; Lemons et al., 2014). In our previous work, for example, we developed a rubric that identified non-negotiables (i.e., essential elements that must be implemented) and negotiables (i.e., aspects where teachers have greater flexibility). This rubric helped teachers make context-sensitive adjustments while ensuring that core components were delivered consistently and with fidelity. As a result, this structured approach supported the successful scaling of the program to a much larger sample, highlighting the value of balancing

program consistency with classroom-level flexibility in future policy and implementation efforts (Neuman et al 2021; Neuman, 2025).

Relatedly, teachers will have to grapple with how best to balance whole-group and small group instruction, particularly during this crucial instructional year. Whole-group instruction can serve a powerful role to systematically and explicitly introduce skills, ensuring that all students of all skill levels have equitable access to grade-level content (Shanahan, 2018). Nevertheless, whole-group lessons have their downside: these lessons cannot adequately differentiate instruction to meet the diverse skill needs present in most classrooms. To help teachers differentiate more effectively, we designed small-group activities grounded in high-leverage practices that adapted the Geodes materials to better match students' needs. For example, we created lessons in phrasing, taking sentences from the current book to help children with chunking; used paragraphs from the book to practice re-reading; adapted versions of the story with shorter sentences for children struggling with comprehension. In this way, teachers could stay within the core materials while providing targeted support. This balance of strong whole-group instruction and purposeful small-group differentiation supports instructional coherence while better addressing diverse learners (Puzio et al., 2020).

It is fair to say that in setting a policy to adopt a mandated program, few could have fully anticipated the day-to-day challenges teachers would face in practice. In fact, subject to a randomized controlled trial at the outside of implementation, one might predict null effects. Design experiments, on the other hand, are pragmatic, as well as theoretical, in the study of both the iterative nature of the design and the resulting ecology of learning (Cobb et al., 2003). This research approach allowed us to surface these unanticipated barriers and work alongside teachers to develop supports that were feasible, relevant, and hopefully, more likely to be sustained. As

our findings indicated, the systematic attention to evidence about learning through progress-monitoring seemed to generate such 'buy-in' among many teachers and classrooms.

This focus on adapting tools and routines in response to real-world conditions underscores the importance of social validity — ensuring that new programs are not only effective in theory, but also feasible, acceptable, and valuable to those responsible for bringing them to life. In doing so, design-based research and a commitment to social validity together may offer a practical safeguard against the pendulum swings that have long challenged the sustained success of reading reforms. By grounding new initiatives in teachers' everyday realities and providing opportunities to refine them in practice, we may be better able to prevent the cycle of rapid adoption and abandonment that has historically undermined progress in reading instruction (Goldenberg, 2020; Wyse & Bradbury, 2022). Rather than swinging from one approach to the next, programs that teachers find usable and worthwhile are more likely to be implemented with fidelity and sustained over time.

To date, however, much of the research on social validity has focused on assessing practitioners' perceptions of an intervention's acceptability only at its conclusion, which limits its potential to inform timely adjustments. In contrast, our study treated social validity as an ongoing tool for gathering actionable feedback throughout implementation. By doing so, especially in the context of adopting a new program in 'uncharted territory,' we were able to refine and adapt our supports in real time to better target improvements in students' reading fluency. This iterative approach highlights the promise of using social validity not merely as an evaluative endpoint, but as a practical mechanism for strengthening the relevance and impact of instructional innovations as they unfold. Notably, when teachers did integrate Geodes more fully

into their instruction, we observed clear improvements in students' accuracy and fluency in reading connected text.

Limitations

There are important limitations in our research. First, we recognize that students' gains in scores cannot be attributed to our supports or necessarily to the result of the materials themselves. Although we would like to believe that helping teachers see the value of the materials and its alignment with the other aspects of the instructional program, such conjectures are merely speculative; further research would require an experimental design. Therefore, we make no claims for causality. Second, much of the evidence for social validity came from teacher self-reports, focus groups and surveys. Such data are inherently subject to perception bias and social desirability effects, which may have led participants to overstate their use of the materials or their comfort with the new instructional routines. At the same time, we would argue that our ongoing observations of teachers' practice helped to mitigate this problem to a great extent. Third, we recognize that concept of social validity and its measurement remain vague, and subject to multiple interpretations. Also, as Larson et al. (2020) note, it has received limited attention in culturally and linguistically diverse settings. More research is needed to clarify the construct and to explore its application across diverse contexts. And finally, our results are context-specific and may not generalize to other contexts. The supports and adaptations that teachers found feasible and worthwhile emerged within a particular district context, with unique coaching structures, local priorities, and student populations. As a result, what worked in this setting may not transfer directly to other schools or districts with different conditions and constraints.

Implications

Taken together, these findings highlight the promise, and in our view, the necessity of treating social validity not as an outcome but as an active driver of implementation. When teachers see how new materials align with their daily practice and students' needs, they are more likely to integrate them with consistency and purpose. This study underscores that fostering such alignment is not a one-time training event, but an ongoing, iterative process that requires listening to teachers' concerns, co-constructing supports, and making visible the connections between materials, instruction, and student learning.

More broadly, our results suggest that large-scale, mandated reforms risk faltering when they ignore the everyday realities of classroom practice. By integrating design-based research principles and centering social validity throughout implementation, districts may be better positioned to sustain reforms beyond their initial rollout. Instead of the familiar pendulum swings that have long characterized reading instruction, policymakers and district leaders could invest in structures that surface practical barriers early on, adapt supports responsively, and build teacher ownership over time.

As districts and states continue to adopt the science of reading as a guiding framework, these insights offer a cautionary reminder: evidence-based programs do not implement themselves. Ensuring that they are perceived as feasible, relevant, and worthwhile by the teachers who bring them to life is essential not only to protect against the pendulum swings of reform but to build the foundations for sustained, meaningful improvement in students' reading achievement.

References

Biggs, E., & Hacker, R. (2021). Engaging stakeholders to improve social validity: Intervention priorities for students with complex communication needs. *Augmentative and Alternative Communication*, *37*, 25-38. https://doi.org/10.1080/07434618.2021.1881824

Birken, S., Haines, E., Hwang, S., Chambers, D., Bunger, A., & Nilsen, P. (2020). Advancing understanding and identifying strategies for sustaining evidence-based practices: a review of reviews. *Implementation Science*, *15*(88). https://doi.org/10.1186/s13012-020-01040-9

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77-101. https://doi.org/10.1191/1478088706qp063oa

Bryk, A., Gomez, L., Grunow, A., & LeMahieu, P. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.

Calkins, L. (2015). *Units of Study-Reading*. Heinemann.

Clayback, K., Williford, A., & Vitiello, V. (2023). Identifying teacher beliefs and experiences associated with curriculum implementation fidelity in early childhood education. *Prevention Science*, *24*, 27-38. https://doi.org/10.1007/s11121-022-01414-z

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*, 9-13.

Collins, T. & Vaughan, C. (2023, November 7). We changed how our NYC school districts teach reading. It's working. https:www.chalkbeat.org/newyork/2023/11/7/23944336/reading-curriculum-literacy-nyc.

Domitrovich, C., Bradshaw, C., Poduska, J., Hoagwood, K., Buckley, J., Olin, S., Romanelli, L., Leaf, P., Greeberg, M., & Ialongo, N. (2008). Maximizing the implementation quality of evidence-based preventive interventions in schools: a conceptual framework. *Advancing School Mental Health Promotion*, *1*, 6-28.

Domitrovich, C., Gest, S., Jones, D., Gill, S., & DeRousie, R. (2010). Implementation quality: Lessons learned in the context of the Head Start REDI trial. *Early Childhood Research Quarterly*, *25*, 284-298.

El Education. (2025). Expeditionary learning. https://www.eleducation.org.

Fixsen, D., Naoom, S., Blase, F., Friedman, R., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*.

Fountas, I., & Pinnell, G. S. (2016). *The Fountas & Pinnell literacy continuum, expanded edition: A tool for assessment, planning, and teaching, preK-grade 8*. Heinemann.

Goldenberg, C. (2020). Reading wars, reading science and English Learners. *Reading Research Quarterly*, *55*(S1), S131-S144. https://doi.org/10.1002/rrq.340

Good, R., Kaminski, R., Cummings, K., Dufour-Martel, C., Petersen, K., Powell-Smith, K., …., & Wallin, J. (2011). *Acadience Reading k-6 Assessment Manual*. Acadience Learning.

Great Minds. (n.d.). Geodes. https://greatminds.org/english/geodes

Great Minds. (n.d.). Wit & Wisdom. https://greatminds.org/english/witwisdom.

Hasbrouck, J., & Michel, D. (2022). *Student-focused coaching*. Paul H. Brookes.

Hasbrouck, J., & Tindal, G. (2017). Oral Reading Fluency Data 2017. https://www.readnaturally.com/assessment-tools

Hiebert, E. (2024). Enhancing opportunities for decoding and knowledge building through beginning texts. *Reading Teacher*, *0*, 1-10. https://doi.org/10.1002/trtr.2303

HMH. (2023). HMH Into reading: K-6 reading curriculum. https://www.hmhco.com/pro

Hoffman, J., Hida, M., & Sailors, M. (2020). Contesting science that silences: Amplifying equity, agency, and design research in literacy teacher prreparation. *Reading Research Quarterly*, *55*(S1), S255-S266. https://doi.org/10.1002/rrq.353

Houghton Mifflin Harcourt. Into Reading. https://www.hmhco.com/programs/into-reading

Joyce, B., & Showers, B. (1996). Research-based professional development. What is it? And what does it look like? *Educational Leadership*, *53*, 5-10.

Kim, J., Burkhauser, M., Quinn, D., Guryan, J., Kingston, H., & Aleman, K. (2017). Effectiveness of structured teacher adaptations to an evidence-based summer literacy program. *Reading Research Quarterly*, *52*, 443-467. https://doi.org/10.1002/rrq.178

Kozleski, E., Hunt, P., Mortier, K., Stepaniuk, I., Fleming, D., Balasubramanian, L., Leu, G., & Munandar, V. (2021). What peers, educators, and principals say: The social validity of inclusive, comprehensive literacy instruction. *Exceptional Children*, *87*, 289-306. https://doi.org/10.1177/0014402920969184

Kraft, M., Blazar, D., & Hogan, D. (2018). The effect of teaching coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*, 547-588. https://doi.org/10.3102/0034654318759268

Larson, A., An, Z., Wood, C., Uchikoshi, Y., Cycyk, L., Hammer, C., Escobar, K., & Roberts, K. (2020). Social validity in early language interventions for dual language learners: A systematic review of the literature. *Topics in Early Childhood Special Education*, *40*, 39-51. https://doi.org/10.1177/0271121419901289

Leko, M. (2014). The value of qualitative methods in social validity research. *Remedial and Special Education*, *35*, 275-286. https://doi.org/10.1177/0741932514524002

Lemons, C., Fuchs, D., Gilbert, J., & Fuchs, L. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher*, *43*, 242-252. https://doi.org/10.3102/0013189X14539189

Lyst, A., Gabriel, S., O'Shaughnessy, T., Meyers, J., & Meyers, B. (2005). Social validity: Perceptions of check and connect with early literacy support. *Journal of School Psychology*, *43*, 197-218. https://doi.org/10.1016/j.jsp.2005.04.004

Mandler, J. M., & Johnson, N. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111-151. https://doi.org/10.1016/0010-0285(77)90006-8

NAEP. (2025). *The Nation's Report Card: Reading Highlights*. Washington DC: National Center for Educational Statistics, Institute of Education Sciences.

Neuman, S.B., Quintero, E. & Reist, K. (2023). *Reading reform across America:A survey of state legislation.* Washington, DC: Al Shanker Institute. http:// https://www.shankerinstitute.org/read.

Neuman, S. B., Samudra, P., & Danielson, K. (2021). Effectiveness of Scaling-up a Vocabulary Intervention for Low-income Children, PreK through First Grade. *Elementary School Journal*, *121*, 385-409. https://doi.org/10.1086/712492

Neuman, S. B. (2025, May 9). Beyond Scripts: Why Structured Adaptations Are Key to Scaling Literacy Program. https://www.shankerinstitute.org/blog/beyond-scripts

New reading curricula for city schools. (2023, June 15). https://www.uft.org/news/news-stories/new-reading-curricula-city schools.

NYC Reads. (n.d.). https://www.schools.nyc.gov/learning/subjects/literacy/nyc-reads

Pearson, P. D. (2004). The reading wars: The politics of reading research and policy—1988 through 2003. *Educational Policy*, *18*, 216-252.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration Policy in Mental Health*, *38*, 65-76. https://doi.org/10.1007/s10488-010-0319-7

Puzio, K., Colby, G., & Algeo-Nichols, D. (2020). Differentiated literacy instruction: Boondoggle or best practice? *Review of Educational Research*, *90*, 459-498. https://doi.org/10.3102/0034654320933536

Rademaker, F., de Boer, A., Kupers, E., & Minnaert, A. (2021). It also takes teachers to tango: Using social validity assessment to refine an intervention design. *International Journal of Educational Research*, *107*(101749). https://doi.org/10.1016/j.ijer/2021.101749

Ratner, N., Han, Y., & Yang, J. (2024). Should we stop using lexical diversity measures in children's language sample analysis? *American Journal of Speech-Language Pathology*, *33*, 1986-2001. https://doi.org/10.1044/2024_AJSLP-23-00457

Schachter, R., Knoche, L., Lu, J., Goldberg, M., Wernick, P., Piasta, S., & Lancaster, H. (2025). A meta-analysis of the effectiveness of coaching and the contribution of coaching processes to learning outcomes for early childhood teachers and children. *Early Childhood Research Quarterly*, *72*, 156-169. https://doi.org/10.1016/j.ecresq.2025.02.014

Snodgrass, M., Chung, M., Meadan, H., & Halle, J. (2018). Social validity in single case research: A systematic literature review of prevalence and application. *Research in Developmental Disabilities*, *74*, 160-173. https://doi.org/10.1177/0741932513490809

The Design-Based Research Collective. (2003). Design-based research: An emergin paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5-8. https://doi.org/10.3102/0013189X032001005

Wheeler, J., & Carter, S. (2019). *The social validity manual*. Academic Press.

Wolf, M. (1978). Social validity: The case for subjective measurement or How applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*, 203-214.

Wyse, D., & Bradbury, A. (2022). Reading wars or reading reconciliation? A critical examination of robust research evidence, curriculum policy, and teachers' pratices for teaching phonics and reading. *Review of Education*, *10*(e3314). https://doi.org/10.1002/rev3.3314

Table 1.

Demographic Characteristics of Teacher and Student Samples

Characteristic	Number
Total number of classrooms	35
Total number of teachers	57
Gender	
Female	53
Male	4
Race/Ethnicity	
White	36
Asian	11
Latina	6
Black	4
Years of teaching	
10+	34
5-9	17
Less than 5	6
Type of Classroom	
NEST (supporting autistic students)	2
G & T (gifted and talented)	2
General Education	18
ICT (integrated co-teaching)	15
Student sample	405
Gender	
Female	50%
Male	50%
Average Age.	6.5 years
Race/ethnicity	
White	49%
Hispanic	16%
Asian	16%
Black	13%
Multi-racial	4%

Table 2.

Summary of Barriers Identified in Surveys and Qualitative Field Notes+

Barrier Theme	Evidence from Post-Workshop Surveys	Evidence from Coaching Field Notes
Time Constraints / Competing Priorities	**78.6%** of teachers reported lack of time as a major barrier to using Geodes	"I'm very overwhelmed with all the things they expect me to implement." "There isn't enough time."
Usability	**42.9%** of teachers felt that the materials did not meet students' skill level	"How do we fit Geodes into our literacy block?" "No one showed us how this fits into what we're already doing."
Requirements of use	**35.7%** of teachers believed that the materials required too much preparation time	"We have Geodes, but no one has used them yet—they remain in boxes." "I'm not even sure where these materials are located."
Skepticism / Questions of Fit	**21.4%** of teachers were not convinced that the materials would improve students' reading skills	"I don't feel it fits my students' needs." "It duplicates things I already do."

+Note: Teachers were asked to select all that applied.

Table 3.

Typology of Geodes Use Across Classrooms

Level	Description of Use	Frequency & Duration	Instructional Context	No. of Classrooms
Level 1: Consistent Use	Geodes used regularly, mostly in small groups. Reinforces phonics, fluency, or comprehension.	3–4x/week; 20–45+ mins total weekly	Whole group intro, primarily small groups, sometimes Independent practice	16
Level 2: Limited-Occasional Use	Geodes used infrequently; seen as optional or supplemental. Usually for struggling readers only.	1–2x/week; 15–30 mins total weekly	Small groups or individual use; less whole group	13
Level 3: Minimal/No Use	Geodes rarely or never used for instruction. Materials kept in library or unused. Barriers include lack of time, perceived misfit, or preference for other materials.	Rare or none	None or unstructured access only	6

Figure 1.

Reading Fluency Gains by Usage Level