



National Reports in Literacy: Building a Scientific Base for Practice and Policy

P. David Pearson and Elfrieda H. Hiebert

The authors examine the National Early Literacy Panel (NELP; 2008) report from two complementary vantage points: (a) the historical tradition of research syntheses in reading research, beginning with Chall and extending through the NELP report, and (b) other recent attempts to examine or synthesize early reading development. While acknowledging the care and precision that characterized the work, the authors of this response raise concerns about the reluctance of the NELP authors to contextualize their findings in relation to both historical and contemporary efforts.

Keywords: early childhood; language processes; literacy

In this article we have combined our independent invitations to provide reviews of the report of the National Early Literacy Panel (NELP; 2008), *Developing Early Literacy* (available at <http://www.nifl.gov/earlychildhood/NELP/NELPreport.html>), for two reasons: (a) because we agree on so many of the virtues and issues of the report and (b) because, by avoiding redundancy between us, we can cover a wider range of issues. Another feature of this response, which may be related to our long personal histories in reading research (Pearson is a decade up on Hiebert, however), is that we take a decidedly historical stance toward this document by reminding readers of the many syntheses of early reading that came before the NELP report and by asking, very pointedly, whether that report adds value to our cumulative knowledge, wisdom, and insight about early reading instruction.

The History

Research syntheses have become an academic art form over the past half-century. By 1960, educational scholars had completely embraced the classic literature review that has been a staple of the field (as evidenced in *Review of Educational Research*, the yearbooks of the National Society for the Study of Education, and the obligatory dissertation chapter that serves as an initiation fee for entry into the academy), and we were starting down the road to what has now become something of an education cottage industry—producing the increasingly ubiquitous *Handbook of X* (where *X* is a variable whose values embrace the entire landscape of educational scholarship). The classic methodology was to empower (or require) a scholar to conduct an exhaustive, critical,

and interpretive review of a given piece of the education research landscape. Gene Glass (1976) made his contribution to this art form by bringing us meta-analysis and the powerful convenience of treating each and every statistical test in each and every experimental study of a phenomenon as a “subject” in a grand experiment as an alternative to the classic literature review synthesis. And like the handbook chapter phenomenon, meta-analysis has become its own independent entity, with an ever-growing literature, its own methodological debates, and a broad programmatic presence in all fields of scholarship (Cooper & Hedges, 1994; Shanahan, 2000).

The Special Case of Reading Research

From time to time, these reviews are blessed with the aegis of professional and/or governmental authority, usually through a commission (sometimes called a panel, task force, or even a committee) that is a part of an effort to promote consensus within fields. And no field has witnessed more synthesis/consensus-seeking efforts than reading, particularly early reading research. For reading, it all began in the 1960s when Jeanne Chall (1967), under the aegis of the Carnegie Corporation, brought us *Learning to Read: The Great Debate* to settle once and for all the question of how to teach beginning reading. Chall made many recommendations, most notable among them a return to an early emphasis on the code; moreover, curriculum developers in the educational publishing industry attended to many of Chall’s recommendations as they developed new programs in the 1970s (Pearson, 1999).

Chall’s (1967) book came out within weeks of the publication of the First-Grade Studies (Bond & Dykstra, 1967), the most ambitious large-scale collaborative effort undertaken by the Cooperative Research Branch of the Office of Education up to that time. And although the First-Grade Studies were not presented as a synthesis, they served that role whether they wanted to or not—precisely because they brought the lens of empirical evaluation to the competing set of “best practices” for teaching early reading. And for a decade or so, things seemed to settle down as phonics made a return to a dominant position in reading pedagogy in the post-Chall years.

The next substantial synthesis came in the middle 1980s when Richard Anderson and his colleagues (Anderson, Hiebert, Scott, & Wilkinson, 1985) offered *Becoming a Nation of Readers* as a National Academy of Education–sponsored response to the frontal assault on educational policy and practice served up by *A*

Nation at Risk (National Commission on Excellence in Education, 1983). It made a strong recommendation for early emphasis on phonics but focused the bulk of its energy on convincing educators and the lay public to embrace the messages of meaning and comprehension that had dominated reading research between the publication of Chall's book and the release of *A Nation at Risk*.

Still (apparently) not content with the synthesis reached in the mid-1980s, the Office of Educational Research and Improvement, under some pressure from congressional mandate (Zorinsky, 1986), directed the Center for the Study of Reading to conduct a more focused synthesis of beginning reading; this led eventually to the publication of Marilyn Adams's (1990) *Beginning to Read: Thinking and Learning About Print*. Adams emerged as a strong code-emphasis proponent but an even stronger proponent for sticking to the evidence. So she hedged on providing an unqualified endorsement of early phonics as the unadorned answer to America's reading problems. She contextualized her endorsement of an early code emphasis by unpacking all of the other rich details and practices that must be present to make a program successful, including lots of reading, writing, and talking about stories and print.

Next in line was the National Academy of Education report *Preventing Reading Difficulties in Young Children* (PRD; Snow, Burns, & Griffith, 1998). PRD was jointly funded and sponsored by the Office of Special Education Programs in the Department of Education, the Office of Educational Research and Improvement (Early Childhood Institute) in the Department of Education, and the National Institute of Child Health and Human Development (NICHD)—Human Learning and Behavior Branch. Like its predecessors, PRD came out squarely for early phonics. However, it added a new element to the early reading mix—an even earlier emphasis (preschool, kindergarten, or early Grade 1) on phonemic awareness (PA) instruction as a prerequisite or corequisite to early phonics instruction. And then, like its predecessors, it went on to suggest that even the most ambitious and effective of early code-emphasis programs must be surrounded and supported by all of the “rich” aspects of oral and written language known to promote the development of comprehension and meaning making. Again, a strong endorsement—but no silver-bullet status—for phonics.

In 1997, Congress asked the director of the NICHD, in consultation with the secretary of education, to convene a national panel to assess the status of research-based knowledge of early reading development, including the effectiveness of various approaches to teaching children to read. This panel, dubbed the NRP (for National Reading Panel), was charged with providing a report that

should present the panel's conclusions, an indication of the readiness for application in the classroom of the results of this research, and, if appropriate, a strategy for rapidly disseminating this information to facilitate effective reading instruction in the schools. If found warranted, the panel should also recommend a plan for additional research regarding early reading development and instruction. (NICHD, 2000, p. 1)

The NRP used the most “scientific” review approaches (i.e., meta-analysis, wherever it could) to distill from existing research

what was known about the efficacy of teaching PA, phonics, fluency (instantiated as either guided reading instruction or independent reading), comprehension, and vocabulary; in addition, the panel investigated the status of the research base on teacher education and professional development and attempted to review research on technology and literacy. The findings from the NRP (NICHD, 2000) were straightforward: Teach PA in K–1, phonics first and fast, comprehension strategies through explicit instruction, vocabulary through a range of approaches, and fluency through oral reading practices. The panel declined to make substantial recommendations about silent reading (claiming the research base was too weak to draw any credible conclusions about its efficacy) and made very modest claims about technology and teacher education. Unlike its ancestral cousins in the synthesis enterprise (save Chall's book), the report of the NRP has proved to be amazingly influential in shaping policy and practice at both the federal level (through the Reading First provisions of No Child Left Behind) and the state level (by virtue of policies designed by states to be aligned with No Child Left Behind—Reading First).

So where does the NELP (2008) report fit in this long tradition of reading research syntheses? The answer to that question constitutes the remainder of this review. We hope to convince readers that the following claims about the report's efficacy and usefulness are warranted by the evidence and arguments we will bring forward:

- The NELP report adds little new knowledge or insight about teaching reading to young children, but it does serve the function, intentional or not, of strengthening the validity of the recommendations emerging from other syntheses, most notably PRD and the NRP report.
- Because the scope of the NELP review did not allow examination of its findings in relation to key contemporary research and evaluation efforts (e.g., the implementation and evaluation of Reading First and Early Reading First and the Early Childhood Longitudinal Study [ECLS]), it does not provide insights or recommendations that can move the field of early literacy instruction ahead. To the contrary, it simply reinforces practices that have already been widely implemented without resounding success.

The Review

NELP Adds Weight and Strength to Earlier Reviews

The findings of NELP focus on two issues: (a) those skills or abilities that, measured early in children's development, predict later literacy proficiency and (b) the effects of interventions (i.e., specific preschool or kindergarten programs, specific instructional emphases in preschool or kindergarten literacy programs such as a code focus or shared reading, home and parent programs, and language enhancement interventions) on supporting those skills or abilities directly. Although the panel provided a chapter for each of five interventions, we believe, for purposes of clarity and brevity, that these programmatic efforts can be clustered together. Regardless of the context (i.e., home, preschool, kindergarten) or content focus (e.g., language, code, shared reading), the studies in each of the reviews described a form of

Table 1
Key Findings From the National Early Literacy Panel Report

Category	Variable	Finding
Predictors	Precursor literacy skills: <ul style="list-style-type: none"> • Alphabet knowledge (knowledge of letters and sounds associated with printed letters) • Phonological awareness • Rapid automatic naming (RAN) of letters or digits • RAN of objects or colors • Writing or writing name • Phonological memory 	Medium to large correlations with later conventional literacy skills
	Additional early literacy skills: <ul style="list-style-type: none"> • Concepts about print • Print knowledge • Reading readiness • Oral language • Visual processing 	Moderate correlations with at least one measure of later literacy achievement
Interventions	Code-focused instruction	Moderate to large effects across a broad spectrum of early literacy outcomes
	Book-sharing interventions	Moderate effects on children’s print knowledge and oral language skills
	Home and parent programs	Moderate to large effects on children’s oral language skills and general cognitive abilities
	Preschool and kindergarten programs Language-enhanced interventions	Moderate to large effects on spelling and reading readiness Large effects on children’s oral language skills

intervention. That is, unlike analyses of the effects of attending *any* kindergarten (e.g., Prince, Hare, & Howard, 2001) or preschool (e.g., V. E. Lee, Brooks-Gunn, Schnur, & Liaw, 1990), the studies in these five chapters of the NELP report analyzed the effects of some form of intervention in preschools, kindergartens, or homes. For convenience in the rest of this review, we summarize the key findings of NELP in Table 1.

Predictors of later literacy proficiency. With respect to early indicators of later success, the panel identified 11 variables that have proved to be moderate to strong predictors of later literacy proficiency. Six of these variables, the panel concluded, served as the “best” (i.e., strongest and most consistent) predictors. Of these 6 variables, 2—alphabet knowledge (which the panel defined to include letter–sound as well as letter–name correspondences) and phonological awareness—proved to be the best of the best. This “double bill” is long-standing in the literature, having been reported as part of federal initiatives for more than 40 years, beginning with the First-Grade Studies (Bond & Dykstra, 1967) and even earlier in the work of Durrell and Murphy (1953). The addition of early writing performance to the list of strong predictors verifies optimistic but ill-substantiated observations made in earlier national summaries (Adams, 1990; Snow et al., 1998); the NELP is the first to offer a substantial database to warrant claims that previously fell into the category of informed expert judgment.

The inclusion of rapid automatic naming (RAN) of all sorts of phenomena (i.e., letters, digits, objects, or colors) on a list of strong predictors is unique to this synthesis. Whether this addition represents a contribution of the report remains to be seen

because, unlike most of the other strong predictors (letter knowledge, PA, writing, and perhaps phonological memory), RAN has not made the transition from “early indicator” to causally related factor through randomized or even natural experiments. For some reason, in the culture of education research and development, we seem unable to resist the temptation to believe that *predictors* are prime candidates to be transformed into *causes* of learning and achievement. To those who develop curriculum, pedagogy, and professional development, we offer the classic cautionary tale, “Don’t confuse correlation with causation,” but we fear that the inclusion of RAN in a list of predictors in a national report—especially when that list appears as the first finding of the report—has potential for egregious, unwarranted, and untested translation to practice. We only hope that we have the collective wisdom to wait for the experiments that can settle the question of their pedagogical significance. This caution seems even more appropriate in light of the weak body of evidence documenting “teaching to” speeded tests of many of the important skills of early reading (Paris, 2005).

Effects of interventions. The review of interventions, from our perspective, represents the strongest potential contribution of the NELP report. But this sort of initiative—synthesizing what we know about various interventions or programmatic elements—is not new, so what is the value added by NELP? We think its value lies in the breadth, depth, and precision of the review. Earlier national syntheses have identified these categories of interventions as useful in developing students’ literacy background and capacity for benefiting from instruction (Adams, 1990; Snow et al., 1998). In the case of the *PRD*, supportive research on book-sharing or oral

language programs was presented as part of a narrative commentary. However, no previous effort has collected *all* of the available evidence on *all* of these programs and examined it through the lens of meta-analysis. It is genuinely useful to know that five programmatic initiatives—namely, code-focused instruction, shared reading interventions, parent and home programs, preschool and kindergarten programs, and language enhancement interventions—all make a consistent difference in profiles of student achievement on outcomes that we value as indicators of reading competence. Even more interesting is that these five general programmatic categories tended to influence different sorts of outcomes, suggesting a kind of “specificity” of effects. For example, language enhancement programs influenced oral language skills but not necessarily code-based knowledge.

This is not to say that the previous efforts did not highlight the benefits of one or more of these types of programs; they did. In fact, in every single synthesis since Chall’s (1967) book, the virtues of early phonics have been solidly proclaimed along with the obligatory caution that phonics should be surrounded by a rich curriculum in other areas. But of these five program types, code-focused instruction is the only one to have achieved the gold standard prior to the NELP report. Thus, on the value-added dimension, the NELP report offers us the possibility of insights that can inform policy and practice by expanding the preschool toolkit available to educators and policy makers.

But alas! After tempting us with the expanded toolkit, NELP panel members leave us hanging—exactly what might be a next step? We’re never told. They have not embraced the mantle of expertise offered by the data—interpreting, asking questions, weighing the contributions of different types of interventions, considering their combined effects; there is no suggestion of what choices might provide the greatest leverage, and not much building on the accumulated body of evidence from earlier synthesis efforts. Granted, the panel provides the facts. But without an authoritative interpretive and comparative lens, the usefulness of the facts is not at all clear. And why are there no illustrative examples from the research to give the reader a more vivid sense of what might matter? We have no idea of what a successful home program looks like. We have no idea of what differentiates a language-enhanced intervention from business-as-usual practices in a preschool classroom.

What makes this failure to assume the mantle of expertise all the more surprising is that this was one of the explicit goals stated in the NELP (2008) report’s introduction:

to synthesize research to contribute to decisions in educational policy and practice that affect early literacy development and to determine how teachers and families could support young children’s language and literacy development. In addition, this evidence would be a key factor in the creation of literacy-specific materials for parents and teachers and staff development for early childhood educators and family-literacy practitioners. (p. iii)

If that was, indeed, panel members’ goal, they have failed themselves and us. And, in the case of code-focused instruction, they have taken a step backward from the immediate predecessor, the NRP (NICHD, 2000).

We base this claim on the comprehensive review conducted by the NRP (NICHD, 2000) on code-focused instruction. The

NELP (2008) states that the NRP did not examine the implications of instructional practices used with children from birth through age 5 (p. v.). In the case of alphabets (NRP’s term to encompass PA and phonics instruction), that observation is simply inaccurate. The subgroup on alphabets defined the scope of its review as “preschoolers, kindergartners, 1st graders, or 2nd through 6th graders” (NICHD, 2000, pp. 2–3). Our point is that there was an existing national database for the NELP. Not only did the NELP not build on the NRP’s analysis, but the NELP’s conclusions fail to go as far as those of the NRP. The generous interpretation of this failure is that the NELP found more to be cautious about than did the NRP; an alternative interpretation is that the NELP failed to exercise its scholarly prerogative in taking the next step in the interpretation of the evidence. The conclusions of the NRP had been quite sophisticated, even nuanced—at the level where useful information is provided to policy makers, practitioners, parents, and publishers (the stated goal of the NELP—p. iii). Not only did the NELP not build on the findings of the NRP, but the NELP’s conclusions fail to extend or even verify the findings of the NRP even though the panel reviewed some of the same studies (a core set of 17). The NRP provided information that, indeed, could be used to guide practice—the stated goal of the NELP. We compare the two sets of recommendations about PA in Table 2.

With respect to PA, the voice of scholars experienced in classrooms and in policy arenas comes through in the NRP in its observations regarding diminishing returns, the need to contextualize PA with letters and to work with young children on particular PA skills consistently, the need to instruct in small groups, the need to start earlier rather than later, and the appropriateness of teachers’ providing instruction. The tenor of these recommendations indicates that the NRP had considered the consequences of potential misinterpretations (e.g., too much, too many, too late).

By contrast, the NELP’s conclusions on PA (Lonigan, Schatschneider, & Westberg, 2008) are ambiguous and generic: Activities should involve higher level PA skills. PA training can occur alone or in combination with print knowledge. Instruction should be either individual or in small groups. There is no point along the developmental continuum when PA instruction is not beneficial. And there is no evidence of teachers’ efficacy at providing PA instruction. Such broad and general findings are ripe for misinterpretations and overextensions in practice. When Congress asks a group of scholars who are content specialists to synthesize research to “contribute to decisions in educational policy and practice,” scholars are obligated to provide more than technical skill at selecting studies and conducting statistical analyses. Technical quality is assumed. The value added of a panel or commission of the field’s most reputable scholars is their ability (and license) to provide the best answers currently available to guide policy and practice. In this task, the NELP took a step back from the answers provided by the NRP, at least with respect to code-based instruction for young readers.

NELP Should Have Examined Contemporary Large-Scale Research Efforts

Surprisingly absent from the NELP report is any mention of an important group of federal initiatives in early childhood literacy,

Table 2
Conclusions of the NRP and the NELP on Phonemic Awareness/Code-Focused Instruction

Issue	National Reading Panel (NRP)	National Early Literacy Panel (NELP)
Activities	“Effect sizes were larger when children received focused and explicit instruction on one or two PA [phonemic awareness] skills than when they were taught a combination of three or more PA skills” (National Institute of Child Health and Human Development [NICHD], 2000, pp. 2–4).	“Interventions should include PA training with activities involving higher-level PA skills, such as actively engaging in analysis or synthesis of words at the syllable, onset-rime, or phoneme level with feedback on correct and incorrect responses” (NELP, 2008, p. 119).
Connection to graphemes	“Instruction that taught phoneme manipulation with letters helped normally developing readers and at-risk readers acquire PA better than PA instruction without letters” (NICHD, 2000, pp. 2–4).	“Some form of PA training, either alone or in combination with more or less complex instruction related to print knowledge (i.e., letter-name instruction, instruction in early decoding skills) is likely to yield growth in children’s skills related to later reading and writing achievement” (NELP, 2008, p. 118).
Size of group	“When children were taught PA in small groups, their learning was greater than when they were taught individually or in classrooms” (NICHD, 2000, pp. 2–4).	“The majority of the code-focused interventions summarized by this meta-analysis were conducted as either individual-level or small group-level interventions. There was no evidence that whole-class or large-group code-focused interventions will produce similar-sized effects on children’s reading-related skills” (NELP, 2008, p. 119).
Length of time	“The length of time spent teaching children was influential, with treatments lasting from 5 to 18 hours producing larger effect sizes than shorter or longer treatments” (NICHD, 2000, pp. 2–4).	“These findings indicate that there is not a point along either an age or a developmental continuum at which code-focused interventions become more or less beneficial to children’s early literacy skills” (NELP, 2008, p. 119).
Individual differences	“Preschoolers exhibited a much larger effect size on reading than did students in the other grade levels. . . . The effects of PA training on reading outcomes were also influenced by SES [socioeconomic status], with mid-to-high SES associated with larger effect sizes than low SES” (NICHD, 2000, pp. 2–4).	“Importantly, there was no evidence that the effectiveness of code-focused interventions was influenced by age or development level of the children. That is, the impacts of code-focused interventions were observed in children whether they were preschool age or kindergarten age, and these interventions were equally successful across a range of levels of prior literacy knowledge (from minimal AK [alphabet knowledge] to being able to read)” (NELP, 2008, pp. 118–119).
Efficacy of classroom teachers	“Classroom teachers were very effective in teaching PA to children” (NICHD, 2000, pp. 2–4).	“The majority of interventions included in these analyses were designed and implemented by researchers” (NELP, 2008, p. 119).

namely, Reading First (which includes kindergarten; Gamse, Jacob, Horst, Boulay & Unlu, 2008), Early Reading First (which targets the same age group as NELP; Jackson et al., 2007), and the ECLS (Denton & West, 2002). Across several careful readings of the NELP report we could not find a single mention of Reading First or of Early Reading First; we did find two, albeit cursory, references to the ECLS. These other reports were clearly available before the official publication of the NELP report. This is particularly the case with reports related to kindergarten from the ECLS (Denton & West, 2002). In the case of the Reading First and Early Reading First reports, the NELP reviewing team had likely completed their meta-analyses before the reports’ appearance. But at the very least the NELP team should have compared their results with these reports in an introductory or ending interpretive perspective. By ignoring these initiatives, the NELP report fails to acknowledge the elephant in the room, namely, that the code-based variables the report identifies as both predictive and causative of success are neither new nor untested;

to the contrary, (a) the evidence for their effectiveness is at least as old as Adams’s *Beginning to Read* volume in 1990, and (b) they *have been* the driving force of the American early childhood literacy curriculum over at least the past decade, perhaps longer. Federal policies and large-scale initiatives have promoted a particular kind of code-centric curriculum in kindergarten through Reading First and for 4-year-olds through Early Reading First. Further, the national kindergarten curriculum has been heavily influenced by the code-driven textbook mandates of California and Texas (California English/Language Arts Committee, 1999; Texas Education Agency, 1997) that almost always determine what is available to the remainder of the country. To assert, as the NELP report does, that the NELP’s definitive findings should point the way to reform represents an implicit denial of the reality of early reading pedagogy in the United States at the present time.

We understand why these national reports were excluded from the main NELP analysis: They do not meet the *prima facie*

Table 3
Percentage of Students in a Cohort Attaining Particular Skills on the Early Childhood Longitudinal Study

Outcome	Fall (Kindergarten)	Spring (Kindergarten)	Spring (Grade 1)
Letter recognition	67	95	100
Beginning sounds	31	74	98
Ending sounds	18	54	94
Sight words	3	14	83
Words in context	1	4	48

Note. From Denton and West, 2002, Figure 2, p. viii.

“peer reviewed journal” test. That is well and good, but it forces the NELP panel to face a curious risk: The panel might include a minor small-*n* study that passed peer review in a third-tier regional journal while excluding a large-*n* federally supported, peer-advised, and well-designed study that was, for reasons of heft, never submitted to a journal. At the very least, the panel might have discussed its findings in relationship to these other national efforts, attempting to explain any similarities and differences in findings or implications for policy.

Recent federal and state initiatives have escalated code-focused instruction in kindergarten (Goldstein, 2007; Hiebert, 2008). In many kindergarten contexts (as well as preschool ones), young children are involved with instruction that aims to promote at least two of the code-based predictor variables identified in the NELP report—letter naming and phonological awareness. In the Executive Summary, Lonigan and Shanahan (2008) state, “Most young children develop few conventional literacy skills before starting school” (p. vii). This is simply not the case, as is evident in the findings of the ECLS (Denton & West, 2002). Although the NELP report does not acknowledge Early Reading First and Reading First, the existence of the ECLS is recognized, albeit in passing. NELP refers to the ECLS report as evidence that there is a substantial amount of variance in conventional literacy within a cohort. About the variance in the ECLS cohort, the panel is right. But there is more to be garnered from ECLS: Differences in the literacy proficiencies of a cohort of young children notwithstanding, the ECLS also indicates that on the strongest predictor variable—alphabet knowledge—the target cohort (class of 2010, since students entered kindergarten in 1998) left kindergarten quite knowledgeable. Data from the ECLS are presented in Table 3. According to the ECLS, 67% of the class of 2010 could recognize letter names at the *beginning* of kindergarten, rising to 95% at the end.

Verification that the ECLS accurately captured the performance of a modern kindergarten cohort comes from the work of Invernizzi, Justice, Landrum, and Booker (2004) when they measured all of the kindergartners in Virginia at the *beginning* of kindergarten. Of the 83,099 children who entered Virginia’s kindergartens in 2003, 17,792 (21.4%) needed intervention based on an average letter-naming score of 5.36 (*SD* = 5.28), whereas the remainder did not need intervention, exhibiting a mean letter-naming performance of 20.85 (*SD* = 4.36). The 79% level of competence in Virginia is higher than the 67% reported by the ECLS, but given that Virginia ranked sixth in the latest state-by-state fourth-grade National Assessment of Educational Progress (NAEP) comparison (75% of its students were at or above basic on NAEP, compared with 65% nationally; J. Lee, Grigg, &

Donahue, 2007), the Virginia–nation discrepancy at kindergarten entry seems comparable.

Is it just an eerie coincidence that the percentage of fourth graders scoring below basic on the NAEP nationally—35% compared with 25% in a state such as Virginia—mirrors almost precisely the percentage of children who have *not* mastered letter naming on entry into kindergarten—33% nationally versus 21% in Virginia? We do not have sufficient evidence—or space in this essay—to explore this connection. So we will simply note the coincidence. The typical argument is that the NAEP is far too distal a measure of achievement to relate to kindergarten entry performance. On the other hand, the consistency with which first graders’ reading achievement predicts achievement in subsequent grades is well established (e.g., Alexander, Entwisle, & Horsey, 1997; Ensminger & Slusarcick, 1992; Juel, 1988).

If kindergarten letter naming is a predictor of first-grade performance, it could be argued that simply getting students to letter naming by the end of kindergarten is not doing the trick. Once students in a cohort have started to read, measures of reading itself become the standard. On reading words in context, the ECLS data (see Table 3) suggest that the cohort of students in the longitudinal sample is not doing particularly well—even though most have alphabet knowledge under control. In other words, relative high competence in alphabet knowledge was not associated with strong word-reading performance. If the relatively high alphabet knowledge scores of this cohort do indeed reflect an increased emphasis of alphabet instruction in prekindergarten and kindergarten, then that instruction seems not to have paid off much in the way of word-reading dividends for these children in kindergarten. This finding suggests that there is little evidence that the literacy trajectories of students who enter school with “below-basic” proficiencies show progress on tasks that really matter as a result of an ever-increasing emphasis on the code.

Could it be that what these knowledgeable kindergartners need is greater attention to word reading and writing rather than to their prerequisites? There is evidence—not included in the NELP report—that appropriate reading instruction in kindergarten can result in positive and meaningful differences in reading proficiency throughout a school career (Hanson & Farrell, 1995). Even in the NELP report, evidence points to facilitative effects of the various interventions (i.e., code focus, book sharing, home/parent, preschool/kindergarten, language focus) on young children’s reading—even when that instruction is not focused directly on reading acquisition. Because the NELP does not engage in interpretive commentary, the panel’s findings on word learning are not considered in depth.

Table 4
Effect Sizes for Measures of Alphabet Knowledge (AK), Phonemic Awareness (PA), and Reading Across Five Types of Interventions

Type of Intervention (Table in NELP)	Measure	Effect Size	n of Studies	p
Code focused (3.1)	AK	.38	24	.0002
	PA	.82	51	< .0001
	Reading	.44	36	< .0001
Shared reading (4.1) ^a	AK	-.06	2	.78
	PA	.11	2	.42
Home and parent (5.1)	AK	-.03	1	.81
	PA	.21	2	.21
	Reading	.28	1	.17
Preschool/kindergarten (6.1)	AK	.23	4	.27
	PA	.08	2	.49
	Reading	.75	9	.19
Language enhancement (7.1)	AK	NA	NA	NA
	PA	.57	2	.05
	Reading	.36	2	.343

Note. From National Early Literacy Panel (NELP), 2008.

^aNone of the studies of shared reading included a measure of reading.

In lieu of this information, we studied the report for data on word reading. Of the five interventions on which the NELP conducted meta-analyses, only shared reading—the very intervention that would be predicted to support reading acquisition—did not include any studies with assessments of reading. Effect sizes for reading, alphabet knowledge, and PA in the various interventions reviewed by the NELP are presented in Table 4. The effect sizes for reading were consistently the highest of any of the three measures. Further, as shown in Table 3.3 of the NELP (2008) report, effects were evident for preschoolers (.75, $p < .01$) as well as kindergartners (.43, $p < .0001$). Students who had little alphabet knowledge showed benefits for reading that were as high as (or higher than) those for alphabet knowledge and PA: reading (.92, $p < .05$), AK (.86, $p < .01$), and PA (.99, $p < .01$; Table 3.4 of the report). In only one of the chapters do the authors refer to the phenomenon of learning to read. Molfese and Westberg (2008), the authors of the NELP chapter on preschool/kindergarten, state the following about the effect size of .75 for reading: “Although, again, this difference did not reach statistical significance, the size of the difference is so large as to be of educational importance. These findings suggest that kindergarten and preschool programs can have an impact on children’s reading development” (p. 199).

We end this section by considering the consequences of ignoring what students know at the beginning and end of kindergarten *and* what they have not learned as a result of efforts that have emphasized the code (the modest if not disappointing findings of the Early Reading First report). Consequential validity has become an increasingly critical consideration in weighing the impact of initiatives in assessment, pedagogy, and policy (Messick, 1989). We need to think of the consequential validity of this

report for all students but especially for the two groups of students who anchor the ends of our distribution of competence—those who are among the 33% who begin kindergarten without conventional literacy (i.e., not knowing letter names or letter-sound matching) and the approximately 31% who begin kindergarten close to mastery of two of the most prominent skills in the typical code-based kindergarten curriculum (i.e., matching initial sounds with letters and recognizing letters).

Consider the first group of students, those in the bottom third of the performance distribution: When this report makes its way into the policy arena, those students will be subjected to an even more aggressive curriculum of “pieces” of language. As Perkins (2008) describes it, they will be faced with “elementitis,” where skills are broken into elements and taught discretely, where playing the whole game of reading is put off until later, once the pieces are in place. They are at risk of falling victim to what we have called the basic-skills conspiracy of good intentions: “First things first, then we’ll get to the good stuff,” so the conspiracy goes. “We won’t dwell long on the code, but surely it must be in place before we get to reading.” Or, “First, let’s make sure they get the letters and words right before we get to the ‘what ifs’ and ‘I wonder whats’ of the curriculum.” This focus on molecular, rather than molar, aspects of the curriculum has surfaced again and again in our history. As far back as 1975, Johnson and Pearson (1975) described the psycholinguistic naïveté of conceptualizing reading as a string of minute behavioral objectives, and as recently as 2005, Paris pointed out the mischief that is done when constrained skills (i.e., finite skills that can be easily mastered, such as alphabet knowledge) prevail over unconstrained skills (i.e., indefinite sets that permit continuous growth such as comprehension or vocabulary breadth). Reading researchers

seem to resist the idea that, somewhere in the reading curriculum, students must learn to orchestrate a wide range of these specific subskills in order to engage in the genuinely skilled practice of reading for meaning.

The more diverse our population becomes, the more we seem to have reified the conventional literacy curriculum, and the NELP report is no exception. However, the evidence suggests that even with high performances on alphabet knowledge (in the data from the ECLS, 95% of a cohort leaves kindergarten with alphabet mastery—see Table 3), students are not doing better on the task of reading as a result of this widespread code-focused curriculum. A long-standing observation that goes back to at least the 1940s (e.g., Gates, 1940) suggests one possible reason: that strong predictors like letter naming are really proxies for a host of other causative variables with which letter-naming scores are naturally correlated—early home or preschool reading experience, book sense, orthographic awareness, even language skill.

Because the NELP report looked only at experimental and predictive research (and excluded largely descriptive work), it tells us nothing about how the students in the top third got to know what they know. A substantial body of evidence supports the view that young children learn a great deal about print and literacy (acquire letter names, grasp the alphabetic principle, and even learn a handful of words) through activities that are fairly typical in some homes but not in others—manipulating magnetic letters, writing words, singing alliterative songs, playing rhyming games, and writing messages to extended family members (e.g., Frijters, Barron, & Brunello, 2000). Call it incidental learning if you like, but asking students to engage in activities that require them to orchestrate many of the components of print literacy (which could be taught and measured separately) does actually result in improvements in these “enabling” skills as well as in students’ capacity to orchestrate them in an integrated performance, which we might even call “reading.” Somehow, in our fascination with all the pieces and predictors, we seem to have lost sight of the goal to which their mastery is linked.

The Likely Legacy of the NELP Report

So, where is the news in NELP? And what will we do differently in our schools as a result of its addition to the long tradition of national syntheses of early reading research? There is some news in NELP, and the big question is how the various strands in the findings will play out in the press and in policy circles. For example, there is some encouraging news for those who would like to see more “meaning-oriented” variables and activities incorporated into the early literacy curriculum. NELP is the first report that has provided gold-standard evidence (i.e., in a meta-analysis) for language predictors, putting them into the second tier in terms of magnitude, just behind the traditional leaders of the pack, alphabet knowledge and PA. And on the intervention side, there is gold-standard evidence to support language and more meaning-based programs, albeit alongside code-focused programs. In fact, the surprising finding about the interventions is how many, not how few, programmatic elements seem to make a difference.

And that brings us to the question of impact and legacy. On both the predictor side and the programmatic side, there is evidence that might lead a policy maker to say, “Well, it’s balance! Both code and meaning predict and shape reading performance.”

But having seen how the NRP was used to rationalize a “first among equals” emphasis on phonics and PA, so that fluency, comprehension, and vocabulary had to wait their turn in the curricular queue, and having seen the persistence of code emphasis in the wake of disappointing evidence about its efficacy in reports on Reading First (Gamse et al., 2008) and Early Reading First (Jackson et al., 2007), we are not sanguine about the capacity of the NELP report to move us toward the kind of balance that its findings would warrant. We wish that the NELP had contextualized its findings both historically and contemporaneously. The historical grounding would have allowed the panel to assess the value added of its report. Had the panel achieved a contemporary grounding, in light of the trends available in the literature on ECLS, Reading First, and Early Reading First, it might have concluded that balance, not focus, is what the research supports. And who knows, the panel might have concluded, as do we, that in moving back to balance as a curricular metaphor, it is high time for the field to reject the pendulum swing, or even the fulcrum, metaphor in favor of an ecological metaphor that argues for, to paraphrase the author of Ecclesiastes (3:1), a time for every curricular purpose under heaven.

REFERENCES

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*(2), 87–107.
- Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Champaign, IL: Center for the Study of Reading.
- Bond, G. L., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly, 2*, 5–142.
- California English/Language Arts Committee. (1999). *English-language arts content standards for California public schools (kindergarten through grade twelve)*. Sacramento: California Department of Education.
- Chall, J. S. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.
- Cooper, H. M., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Denton, K., & West, J. (2002). *Children’s reading and mathematics achievement in kindergarten and first grade*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Durrell, D. D., & Murphy, H. A. (1953). The auditory discrimination factor in readiness and reading ability. *Education, 73*, 556–560.
- Ensminger, M. E., & Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education, 65*(2), 95–113.
- Frijters, J. C., Barron, R. W., & Brunello, M. (2000). Direct and mediated influences of home literacy and literacy interest on prereaders’ oral vocabulary and early written language skill. *Journal of Educational Psychology, 92*, 466–477.
- Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First Impact Study final report*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Gates, A. I. (1940). A further evaluation of reading-readiness tests. *Elementary School Journal, 40*(8), 577–591.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.

- Goldstein, L. S. (2007). Beyond the DAP versus standards dilemma: Examining the unforgiving complexity of kindergarten teaching in the United States. *Early Childhood Research Quarterly*, 22(1), 39–54.
- Hanson, R. A., & Farrell, D. (1995). The long-term effects on high school seniors of learning to read in kindergarten. *Reading Research Quarterly*, 30, 908–933.
- Hiebert, E. H. (2008). The (mis)match between texts and students who depend on schools to become literate. In E. H. Hiebert & M. Sailors (Eds.), *Finding the right texts for beginning and struggling readers: Research-based solutions* (pp. 1–21). New York: Guilford.
- Invernizzi, M., Justice, L., Landrum, T., & Booker, K. (2004). Early literacy screening in kindergarten: Widespread implementation in Virginia. *Journal of Literacy Research*, 36(4), 479–500.
- Jackson, R., McCoy, A., Pistorino, C., Wilkinson, A., Burghardt, J., Clark, M., et al. (2007). *National evaluation of Early Reading First: Final report*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.
- Johnson, D. D., & Pearson, P. D. (1975). Skills management systems: A critique. *Reading Teacher*, 28(8), 757–764.
- Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grades. *Journal of Educational Psychology*, 80, 437–447.
- Lee, J., Grigg, W. S., & Donahue, P. L. (2007). *The nation's report card: Reading 2007* (NCES 2007–496). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Lee, V. E., Brooks-Gunn, J., Schnur, E., & Liaw, F. (1990). Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development*, 61(2), 495–507.
- Lonigan, C. J., Schatschneider, C., & Westberg, L. (2008). Impact of code-focused interventions on young children's early literacy skills. In National Early Literacy Panel, *Developing early literacy: Report of the National Early Literacy Panel* (pp. 107–151). Washington, DC: National Institute for Literacy. Available at <http://www.nifl.gov/earlychildhood/NELP/NELPreport.html>
- Lonigan, C. J., & Shanahan, T. (2008). Executive summary of the report of the National Early Literacy Panel. In National Early Literacy Panel, *Developing early literacy: Report of the National Early Literacy Panel* (pp. v–xii). Washington, DC: National Institute for Literacy. Available at <http://www.nifl.gov/earlychildhood/NELP/NELPreport.html>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan and the American Council on Education.
- Molfese, V., & Westberg, L. (2008). Impact of preschool and kindergarten programs on young children's early literacy skills. In National Early Literacy Panel, *Developing early literacy: Report of the National Early Literacy Panel* (pp. 189–209). Washington, DC: National Institute for Literacy. Available at <http://www.nifl.gov/earlychildhood/NELP/NELPreport.html>
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy. Available at <http://www.nifl.gov/earlychildhood/NELP/NELPreport.html>
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Pub. No. 00–4769). Washington, DC: U.S. Government Printing Office.
- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, 40(2), 184–202.
- Pearson, P. D. (1999). Essay book reviews: A historically based review of *Preventing Reading Difficulties in Young Children*. *Reading Research Quarterly*, 2(34), 231–246.
- Perkins, D. N. (2008). *Making learning whole: How seven principles of teaching can transform education*. San Francisco: Jossey-Bass.
- Prince, D. L., Hare, R. D., & Howard, E. M. (2001). Longitudinal effects of kindergarten. *Journal of Research in Childhood Education*, 16(1), 15–27.
- Shanahan, T. (2000). Research synthesis: Making sense of the accumulation of knowledge in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 209–226). Mahwah, NJ: Lawrence Erlbaum.
- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Texas Education Agency. (1997). *Proclamation of the State Board of Education advertising for bids on textbooks*. Austin, TX: Author.
- Zorinsky, E. (1986). *Amendment #2202 to the Human Services Reauthorization Act (Public Law 99–425)*. Washington, DC: U.S. Congress.

AUTHORS

P. DAVID PEARSON is a professor in the areas of language, literacy, and culture, and cognition and development, in the Graduate School of Education at the University of California, Berkeley, 1501 Tolman Hall, Berkeley, CA 94720; ppearson@berkeley.edu. His research focuses on reading processes, pedagogy, policy, and assessment.

ELFRIEDA H. HIEBERT is an adjunct professor in cognition and development in the Graduate School of Education at the University of California, Berkeley, 4425 Tolman Hall, UC Berkeley, Berkeley, CA 94720; hiebert@berkeley.edu. Her research emphasizes early reading instruction, particularly the critical role played by the texts that young readers are given to learn to read.

Manuscript received October 22, 2009

Revisions received February 4, 2010,

and February 12, 2010

Accepted March 16, 2010