Preschool Through Third Grade Alignment and Differentiated Instruction: A Literature Review

Preschool Through Third Grade Alignment and Differentiated Instruction: A Literature Review

August 2016

Prepared for:

Policy and Program Studies Service
Office of Planning, Evaluation and Policy Development
U.S. Department of Education

Prepared by:

Katie Drummond
Aleksandra Holod
Marie Perrot
Antonia Wang
American Institutes for Research
Washington, DC 20007

Michèle Muñoz-Miller
Mackson Ncube
Herb Turner
Analytica
Phoenixville, PA 19460

This report was produced under U.S. Department of Education Contract No. EDPEP-11-O-0089 with American Institues for Research. Erica Lee served as the contracting officer's representative. This report contains information which references standards, curriculum, instruction and/or assessments. The U.S. Department of Education does not support any particular standard, curriculum, instruction or assessment. The reerences are for informational purposes only and for convenience to the reader. The opinions expressed herein do not necessarily represent the positions or policies of the Department of Education. The inclusion of any resources shall not be construed to interpreted as an endorsement by the U.S. Department of Education of any organization or its business, products or services referenced herin.

U.S. Department of Education
John B. King, Jr.
Secretary

Office of Planning, Evaluation and Policy Development
Amy McIntosh
Delegated Duties of Assistant Secretary

Policy and Program Studies Service
Jennifer Bell-Ellwanger
Director

August 2016

This report is available on the Department's website at:
http://www2.ed.gov/about/offices/list/opepd/ppss/reports.html

Availability of Alternate Formats
Requests for documents in alternate formats such as Braille or large print should be submitted to the Alternate Format Center by calling 202-260-0852 or by contacting the 504 coordinator via email at om_eeos@ed.gov.

Notice to Limited English Proficient Persons
If you have difficulty understanding English, you may request language assistance services for Department information that is available to the public. These language assistance services are available free of charge. If you need more information about interpretation or translation services, please call 1-800-USA-LEARN (1-800-872-5327) (TTY: 1-800-437-0833), or email us at: Ed.Language.Assistance@ed.gov. Or write to: U.S. Department of Education, Information Resource Center, LBJ Education Building, 400 Maryland Ave. SW, Washington, DC 20202.

Content Contact:
Erica Lee
Phone: 202-260-1463
Email: erica.lee@ed.gov

Contents

List of Exhibits

Acknowledgments

We wish to thank several individuals who contributed to the completion of this literature review.

We are grateful to AIR reviewers Eboni Howard and Kerstin LeFloch. As part of the study's technical working group, we also thank Esther Buch, Margaret Burchinal, and Ellen Kisker for their review of the document, and Linda Espinosa and Kristie Kauerz for their input on the topics and design.

Although we appreciate the assistance and support of the individuals listed above, any errors in judgment or fact are, of course, the responsibility of the authors.

Executive Summary

This literature review provides a review of policies, programs, and practices that have the potential to help students sustain the positive effects of preschool as they progress from kindergarten through grade 3 (K–3). The U.S. Department of Education's Policy and Program Studies Service commissioned this systematic literature review, which focuses on two specific approaches: (1) preschool and K–3 alignment, and (2) differentiated instruction in kindergarten and first grade.

Background

Research shows that participation in a high-quality preschool can improve young children's readiness skills for elementary school, positively influencing behavioral, social-emotional, and cognitive outcomes (Andrews, Jargowsky, & Kuhne, 2012). Specifically, for children who may be at risk for academic challenges in early elementary school, attending a high-quality preschool can improve test scores and attendance, and it can reduce grade-level retention and placement in special education (Andrews et al., 2012; Barnett, 2008; Karoly & Bigelow, 2005; Reynolds, 1993; Reynolds et al., 2007). However, some preschool program evaluations document that strong initial benefits may not persist into early elementary school (Lipsey, Farran, & Hofer, 2015; Magnuson, Meyers, Ruhm, & Waldfogel, 2005; Manship, Madsen, Mezzanotte, & Fain, 2013; Ramey et al., 2000; U.S. Department of Health and Human Services, 2010).

Preschool benefits may not persist for many reasons, including lack of continuous follow-up with participating students, lack of family supports or involvement, or limited intensity or duration of the preschool program (Brooks-Gunn, 2003; Halpern, 2013; Reynolds, Magnuson, & Ou, 2006). The positive effects of preschool may not persist if children attend poor-quality elementary schools after preschool (Clements, Reynolds, & Hickey, 2004; Lee & Loeb, 1995). Without additional and continuous supports as children continue through the early elementary grades, participation in preschool cannot overcome potential challenges that children, particularly those at risk for poorer academic outcomes, may face. It is important to identify ways to sustain early cognitive, social-emotional, and academic gains in order to give all students opportunities to thrive academically. To explore potential ways to sustain the positive effects of preschool, this literature review focused on two specific topics: (1) preschool and K–3 alignment and (2) differentiated instruction in kindergarten and first grade. The U.S. Department of Education's Policy and Program Studies Service (PPSS), in collaboration with the Office of Early Learning, selected eight topics for preliminary searches after initial attempts to identify interventions specifically designed to sustain the benefits of preschool turned up low yields. Based on the search results (and after receiving input from multiple Department offices), PPSS recommended two final topics for the literature review. PPSS made final decisions about further specifications for the differentiated instruction section (e.g., only include research spanning grades K–1 and exclude studies that focus exclusively on lower-achieving students).

Preschool and K–3 Alignment

The first topic focuses on approaches to align preschool and kindergarten through grade 3. Preschool or prekindergarten and K–3 alignment (sometimes called *P–3*) emphasizes

coordination among standards, curricula, instructional practices, student assessment, and teacher professional development between the preschool years and the early elementary school years. Early childhood experts assert that the effects of preschool may be sustained and investment in early education capitalized upon if curricula and instructional strategies from preschool through grade 3 are well aligned (Bogard & Takanishi, 2005; Brooks-Gunn, 2003; Howard, 2008). As Reynolds and Temple (2008) suggest, P–3 programs may provide more continuity and better organization of services for students as well as enhanced school-family partnerships.

Differentiated Instruction

The second topic focuses on differentiated instruction in kindergarten and first grade. The premise of differentiated instruction is that teaching practices and curricula should vary to meet the diverse needs and skills of the individual student and to optimize students' learning experiences (Tomlinson, 2000, 2001). In a differentiated instructional delivery model, student needs are emphasized (Stanford & Reeves, 2009), with teachers purposively adapting instructional strategies and the focus of skill building to be responsive to individual or groups of students (Jones, Yssel, & Grant, 2012). One explanation for why initial benefits of preschool do not persist as students enter elementary school is that children who make early gains in preschool may not have the opportunity to maintain their growth rate or learning trajectory because early elementary instruction may focus on students who are less prepared and have low-level skills. In other words, instruction may not be differentiated, and in some cases may not be rigorous enough, to meet and build upon the skills that some students have upon school entry (Claessens, Engel, & Curran, 2013; Kauerz, 2006; Lipsey, Farran, & Hofer, 2015).

For this review, studies were limited to those that involve students in kindergarten or first grade. Because the justification for this topic involves the use of differentiation to meet the skill levels of children *upon their entry to elementary school*, studies that focused exclusively on grades beyond kindergarten and first grade were excluded. Studies that included older grades (i.e., second and third grades) in addition to the earlier grades were retained. The review also excluded studies that focused exclusively on low-achieving students because of the priority on differentiated instruction as a way to help sustain the gains children make in preschool. Studies that include a spectrum of achievement levels (lower achievement in addition to typical or higher achievement) were retained. Finally, although differentiated instruction is consistent with response to intervention (RTI) models and multi-tiered systems of prevention or support (Gettinger & Stoiber, 2012), for the purposes of this review, the focus was on individualization of instruction that takes place within the regular classroom. This review focused only on interventions conducted by teachers in the classroom and not on RTI models as a whole.

Questions

1. What approaches does the research and theoretical literature suggest for aligning preschool through third-grade (P−3) education, and what is the quality of the research studies?

2. What are the findings from studies of differentiated instruction for children in kindergarten and first grade, and what is the quality of these studies?

Literature Review Methodology

To gather appropriate literature, the review team conducted keyword searches related to the two topic areas in nine widely used education and psychology electronic databases. Additionally, for P–3 alignment, the research team determined that articles on the topic may not be widely published in education and psychology journals. For this reason, the research team used additional Internet searches, and requests to experts in the field, including our technical working group members, for article or intervention recommendations. For both topics, articles needed to be published between January 2003 and July 2014 and interventions needed to take place in the United States (including U.S. territories and tribal areas). Because preliminary searches revealed there would be few experimental or quasi-experimental studies for either topic, the research team conducted a broad review to catalog all available studies, and quantify and categorize the currently available research (Brett, Staniszewska, Newburn, Jones, & Taylor, 2011; EPPI Centre, 2010).

All studies that used quantitative designs—including randomized controlled trials (RCTs), quasi-experimental designs (QEDs), and pre-test/post-test and correlational designs—were included if they focused on child-level developmental outcomes, such as academic outcomes (i.e., literacy, mathematics, science), cognitive outcomes (e.g., IQ, language), and/or social and behavioral outcomes for students (e.g., social-emotional, executive functioning). Child outcomes could be measured by standardized achievement tests, researcher- or teacher-developed assessments, post-intervention class grades, student promotion to the next grade, or other measurement approaches. Studies that used primarily qualitative methods were included if they focused on implementation issues relevant to interventions for either topic. Most often, the qualitative studies were case studies—that is, research that seeks close examination of a single program to provide readers with a practical example and/or unique explanations of phenomena (e.g., Hays, 2004).

For preschool and K–3 alignment, as it became clear that the literature did not contain many data-based studies (and no experimental or quasi-experimental designs), the research team decided to include articles in this literature review that cover the theory supporting P–3 alignment and/or policy considerations relating to P–3 alignment.

For differentiated instruction, a substantial number of data-based studies emerged related to the topic. Therefore, theory and policy articles were not included in this literature review. For the subset of quantitative studies that employed a rigorous design, namely an RCT or QED, the research team appraised the research methods to provide more information about the quality of available evidence. The team used the systematic research standards in the What Works Clearinghouse (WWC)™ Single Study Review Protocol (WWC, 2010b) to guide its coding. These standards relate to the amount of confidence that can be placed in a study to demonstrate causal evidence and, subsequently, if a study meets standards, to evaluate the effectiveness of the intervention itself.

Preschool and K–3 Alignment Findings

The P–3 alignment topic includes 49 policy or theory resources, nine qualitative studies, three quantitative studies, and one mixed-methods study. None of the quantitative studies used experimental or quasi-experimental designs to examine impacts of preschool and K–3 alignment

interventions. Reflecting the state of the research in the field, key findings for preschool and K–3 alignment focus on theoretical and policy considerations.

- Nearly all qualitative studies and policy and theory articles on P–3 alignment suggest aligning standards, curriculum, instruction, assessments, and environments across preschool and grades K–3.
- Numerous policy articles call for more similar teacher education and training requirements across preschool and elementary education job positions, and several qualitative studies provide examples of this practice. Authors suggest that preschool teachers should earn bachelor's degrees, hold certification, and receive compensation on par with elementary teachers and that K–3 elementary school teachers should receive more training in early childhood development.
- Numerous policy articles recommend the creation of systems that link individual student data from public and private early childhood programs, particularly preschool programs, to students' public school data so that elementary teachers have more complete and accessible information about students' learning trajectories. With access to these data, educators could better tailor instruction to meet students' needs.
- Several policy articles and several qualitative studies suggest that school district administrators can support the implementation of P−3 initiatives through the management practices they put in place. Specific leadership considerations include the following: (1) involving early childhood education providers and grade K–3 teachers in planning P–3 initiatives, (2) implementing the planned elements of P–3 initiatives with fidelity, (3) specifying measurable student achievement benchmarks, and (4) holding principals and teachers accountable for achieving benchmarks. Two study authors also link similar principal management practices to implementation of P–3 initiatives.
- Several challenges must be addressed if P–3 initiatives are to be more widely implemented, according to the policy literature. A number of qualitative studies illustrate these challenges, which include the following: (1) policies that inhibit the blending of federal, state, and local sources of funding to support P−3 initiatives; (2) instability of preschool funding; (3) resistance by practitioners to integration of preschool and the K–3 grades; and (4) the organization of elementary education classrooms, buildings, and enrollment.

Differentiated Instruction Findings

The differentiated instruction topic includes 21 studies, including 17 quantitative studies and 4 qualitative studies focused on students in kindergarten or grade 1. Of the 17 quantitative studies, 7 were RCTs, 6 were QEDs, and 4 were other non-rigorous designs (i.e., descriptive and single-group pre-test/post-test designs) to examine the effects of differentiated instruction on achievement. Nearly all quantitative studies had methodological issues that diminish the level of confidence in the study to demonstrate causal evidence of effectiveness. Of the 21 studies, most focused on reading instruction (14). Three studies evaluated differentiated instruction on writing outcomes. Four studies examined implementation of differentiated instruction in mathematics. The key findings summarize the results of all reviewed studies, regardless of the study design or the strength of the evidence.

- Of the 17 quantitative studies of differentiated instruction, one RCT of the *Individualized Student Instruction With Assessment to Instruction* intervention demonstrated positive results on reading outcomes and had the potential to meet the criteria for strong causal evidence. Five RCTs of this specific intervention that did not meet the criteria for strong causal evidence also showed positive outcomes.
- One RCT compared the strategies of (1) grouping students by learning style preferences (i.e., visual, auditory, tactile, or kinesthetic), with (2) grouping students by pre-intervention reading achievement. There were no discernible effects in favor of grouping method. This study had a methodological issue because the reliability and validity of the outcome measure was unclear.
- Seven other quantitative studies examined small-group differentiated instruction approaches for reading and showed mixed results. Among these seven (five QEDS, one pre-test/post-test design, and one descriptive design), none meet all criteria designed to evaluate whether a study strongly demonstrates causal evidence, either because of their research designs or because of methodological issues within the designs.
- Three other quantitative studies suggest that some students may benefit from collaborative, interactive writing sessions or from specific writing tools or prompts. The three studies included one QED that failed to appropriately demonstrate baseline equivalence and two single-group pre-test/post-test design studies that cannot show causal evidence of effectiveness due to the research design.
- In addition to the quantitative studies, four qualitative studies provided information about processes and strategies for implementing differentiated instruction for mathematics but do not provide evidence of effects. These small studies, which focused on perceptions of facilitators or barriers to implementation, suggest that differentiated instruction requires careful planning and reflection on the part of teachers. Opportunities for peer collaboration and guidance by mentors, such as coaches, may be helpful to improve teacher practice related to differentiation.

I. Introduction

Research shows that participation in high-quality preschool can improve young children's readiness skills for elementary school, positively impacting behavioral, social-emotional, and cognitive outcomes (Andrews, Jargowsky, & Kuhne, 2012). Specifically, for children who may be at risk for academic challenges in early elementary school, attending a high-quality preschool can improve test scores and attendance and reduce placement in special education and grade-level retention (Andrews et al., 2012; Barnett, 2008; Karoly & Bigelow, 2005; Reynolds, 1993; Reynolds et al., 2007). Studies have demonstrated that high-quality early education is related to other positive developmental outcomes for children, including improved language development, cognitive functioning, social competence, and emotional adjustment (Clarke-Stewart, Vandell, Burchinal, O'Brien, & McCartney, 2002; Howes, 1988; National Institute of Child Health and Human Development Early Child Care Research Network, 2000; Peisner-Feinberg et al., 2001). Additional long-term benefits of attending a high-quality preschool program include higher rates of high school completion, a greater likelihood of attending college, and increased lifetime earnings (Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010; Karoly, Kilburn, & Cannon, 2005; Reynolds & Ou, 2011; Reynolds & Temple, 2008).

Because of the importance of early childhood education, the federal government supports preschool education through the U.S. Department of Health and Human Services' (HHS) Head Start program; through the U.S. Department of Education's (the Department's) special education preschool program, authorized through the *Individuals with Disabilities Education Act*, Part B; and through the new Department- and HHS-administered Preschool Development Grant program. States and local districts also have implemented public preschool programs, many of which are targeted to disadvantaged children and are showing positive results (see Frede, Jung, Barnett, Lamy, & Figueras [2007], Gilliam & Zigler [2001], and Gormley & Phillips [2005] on Oklahoma's universal preschool program in Tulsa, and Weiland & Yoshikawa [2013] on Boston's public preschool).

Importantly, research also shows that not all students who experience preschool achieve positive, long-term outcomes (Barnett, 2008; Lee & Loeb, 1995). Some preschool program evaluations document that strong initial benefits do not persist into early elementary school (Lipsey, Farran, & Hofer, 2015; Magnuson, Meyers, Ruhm, & Waldfogel, 2005; Manship, Madsen, Mezzanotte, & Fain, 2013; Ramey et al., 2000; U.S. Department of Health and Human Services, 2010). Preschool benefits may not persist for many reasons, including lack of continuous follow-up with participating students, lack of family supports or involvement, or limited intensity or duration of the program (Brooks-Gunn, 2003; Halpern, 2013; Reynolds, Magnuson, & Ou, 2006). The positive effects of preschool may not be sustained if children attend poor-quality elementary schools after preschool (Clements, Reynolds, & Hickey, 2004; Lee & Loeb, 1995). Without additional and continuous supports as children continue through the early elementary grades, participation in preschool cannot overcome potential challenges that children, particularly those at risk for poorer academic outcomes, may face. It is important to identify ways to sustain early cognitive, social-emotional, and academic gains in order to give all students opportunities to thrive academically.

Overview

To better understand how to build on the positive effects of preschool, the Department's Policy and Program Studies Service initiated a literature review, consisting of two components:

Part 1: A systematic literature review of policies, programs, and practices that have the potential to aid practitioners and policymakers in helping students in kindergarten through grade 3 (K–3) build on the positive effects of preschool and make cognitive, social-emotional, and academic gains. This review focuses on two questions:

1. What approaches does the research and theoretical literature suggest for aligning preschool through third-grade (P−3) education, and what is the quality of the research studies?

2. What are the findings from studies of differentiated instruction for children in kindergarten and first grade, and what is the quality of these studies?

Part 2: Case study descriptions of five programs that help disadvantaged students in K–3 have positive cognitive, social-emotional, and/or academic outcomes and may build on the positive effects of preschool by using policies, programs, and practices from the two topic areas above. Research questions include the following:

3. What are the characteristics (e.g., resources, personnel, staff characteristics, training, setting, population served) of P–3 or differentiated instruction programs that aim to increase cognitive, social-emotional, or academic outcomes of students?

4. On what research, theory, and/or experiences did the designers of these programs base the program structure and content?

5. What are the challenges of implementing these programs, and how have staff and leaders tried to overcome these challenges?

6. How does the organization implementing the program ensure its sustainability?

The Department selected these topics as the focus of the literature review after preliminary literature searches revealed that there would be few results for the broader topic of the Request for Task Order ("interventions to sustain effects of preschool"). This report includes findings from the literature review and answers to the first two questions. The Department expects to release findings from the case studies in late 2016.

Literature Review Methodology

Various types of systematic reviews can be used to examine extant research literature on particular interventions or approaches to answer questions ranging from "What research exists?" to "What interventions work?" (see Cooper, 2010; EPPI Centre, 2010; Petticrew & Roberts, 2006; What Works Clearinghouse [WWC™], 2010a). The current review balanced these two questions. Because preliminary searches revealed there would be few experimental or quasi-experimental studies for either topic, the research team conducted a broad review to catalog all

available studies to quantify and categorize the currently available research (Brett, Staniszewska, Newburn, Jones, & Taylor, 2011; EPPI Centre, 2010).

Literature Criteria and Search Process

To gather appropriate literature, the review team conducted keyword searches relevant to the two topic areas in nine widely used education and psychology electronic databases (see Appendix A for details on keywords and databases). Searches focused on articles published between January 2003 and July 2014, with approaches taking place in the United States (including U.S. territories and tribal areas).

Preschool and K–3 Alignment

The first topic focuses on approaches to align preschool and kindergarten through grade 3. Preschool or prekindergarten and K–3 alignment (sometimes called *P–3*) emphasizes coordination among standards, curricula, instructional practices, student assessment, and teacher professional development between the preschool years and the early elementary school years. Early childhood experts assert that the effects of preschool may be sustained and investment in early education capitalized upon if curricula and instructional strategies from preschool through grade 3 are well aligned (Bogard & Takanishi, 2005; Brooks-Gunn, 2003; Howard, 2008). As Reynolds and Temple (2008) suggest, P–3 programs may provide more continuity and better organization of services for students as well as enhanced school-family partnerships. Policy authors also suggest that P–3 approach may be particularly beneficial to close achievement gaps for low-income students, English learners, and students with behavior problems (Demanchick, Peabody, & Johnson, 2009; Garland, 2011; Jacobson, 2009; Rice 2008a; Severns, 2012).

Based on the preliminary searches conducted in preparation for the literature review and consultation with a technical working group that advised on the literature review, we determined that articles on P–3 alignment are not widely published in education and psychology journals and therefore do not appear frequently in traditional database searches. For this reason, the research team used additional search approaches, including examination of topic-specific websites (e.g., Foundation for Child Development), general Internet searches, and requests to experts in the field, including our technical working group members, for article or intervention recommendations. Appendix B contains references included in the P–3 review.

Differentiated Instruction

The second review topic includes research studies that focus on differentiated instruction. The premise of differentiated instruction is that teaching practices and curricula should vary to meet the diverse needs and skills of the individual student and to optimize students' learning experiences (Tomlinson, 2000, 2001). It moves away from a one-size-fits-all approach to teaching and from the expectation that learners, themselves, must adapt to preexisting strategies or a set level of instruction. Instead, in a differentiated instructional delivery model, student needs are emphasized (Stanford & Reeves, 2009), with teachers purposively adapting instructional strategies and the focus of skill building to be responsive to individual or groups of students (Jones, Yssel, & Grant, 2012). Some experts assert that differentiated instruction differs from typical ability grouping because teachers maintain high expectations for all students but

respond to student differences in their teaching (Bofferding, Kemmerle, & Murata, 2012; Murata, 2013).

One explanation for why effects of preschool could diminish in early elementary school is that children who make early gains in preschool may not have the opportunity to maintain their growth rate or learning trajectory because early elementary instruction may focus on students who are less prepared and have lower-level skills. In other words, instruction is not differentiated and, in some cases, may not be rigorous enough, to meet and build on the skills that some students have upon school entry (Claessens, Engel, & Curran, 2013; Kauerz, 2006).

In addition to the basic search criteria related to the overall topic, outcome, year of publication and location of the intervention, we applied several additional parameters to the differentiated instruction studies.

- First, we only retained studies that focused on differentiated instruction *interventions*, defined as (1) comprehensive or supplemental instructional *programs* or (2) clearly defined and described *practices.*
- Second, studies were limited to those that involve students in kindergarten and/or first grade. Because the justification for this topic involves the use of differentiation to meet the skill levels of children *upon their entry to elementary school*, studies that focused exclusively on grades beyond kindergarten and first grade were excluded. Studies that included older grades (i.e., second and third grades) in addition to the earlier grades were retained.
- Third, the review excluded studies that focused exclusively on lower-achieving students. They were excluded because justification for this topic involves the use of differentiation to build upon existing skills (potentially attained earlier in preschool). Studies that include a spectrum of achievement levels (lower achievement in addition to typical or higher achievement) were retained.
- Finally, although differentiated instruction is consistent with response to intervention (RTI) models and multi-tiered systems of prevention or support (Gettinger & Stoiber, 2012), for the purposes of this review, the focus was on individualization of instruction that takes place within the regular classroom. In general, RTI models aim to (1) screen students to document their skill levels, (2) deliver evidence-based instruction, (3) monitor students' continued progress, and (4) adjust instruction based on that monitoring (Metcalf, 2013). RTI models could include supplemental, pull-out instruction as educators provide support to students who struggle with skill development. This review focused only on interventions conducted by classroom teachers in the classroom and not on RTI models as a whole. At least 50 percent of the students needed to be general education students; we excluded studies that focused more exclusively on special education.

Appendix C contains references included in the differentiated instruction review.

Study Types

All studies that used quantitative designs—including randomized controlled trials (RCTs), quasi-experimental studies (QEDs) and pre-test/post-test and correlational designs—were included if they focused on child-level developmental outcomes, such as academic outcomes (i.e., literacy, mathematics, science), cognitive outcomes (e.g., IQ, language), and/or social and behavioral outcomes for students (e.g., social-emotional, executive functioning). Child outcomes could be measured by standardized achievement tests, researcher- or teacher-developed assessments, post-intervention class grades, student promotion to the next grade, or other measurement approaches. Studies that used primarily qualitative methods were included if they focused on implementation issues relevant to interventions for either topic or on the outcomes named previously. Most often, the qualitative studies were case studies—that is, research that seeks close examination of a single program to provide readers with a practical example and/or unique explanations of phenomena (e.g., Hays, 2004).

For preschool and K–3 alignment, as it became clear that the literature did not contain many data-based studies (and no experimental or quasi-experimental designs), the research team decided to include articles in this literature review that cover the theory supporting P–3 alignment and/or policy considerations relating to P–3 alignment. For differentiated instruction, a substantial number of data-based studies emerged related to the topic. Therefore, theory and policy articles were not included in the literature review for the differentiated instruction topic.

The nature of the case studies was quite different for the two topics. For P–3 alignment, the case studies focused on implementation of P–3 approaches in a specific state or district. The researchers tended to collect implementation data from various sources, including interviews with stakeholders (e.g., superintendent, board members, principals, teachers, parents), observations of classrooms, and extant documents. For differentiated instruction, the case studies were most often reports from a single school or a small set of classrooms (sometimes one classroom) that had implemented a differentiated instructional strategy. These studies tended to take a practitioner research approach (Pritchard, 2002), also called teacher research or practitioner inquiry (Cochran-Smith & Lytle, 1999a), in which teachers document their own practice. As Ravitch (2014) explains, in an effort to improve practice and influence policy, practitioner research involves practitioners making structured inquiries about aspects of their practice for which they have questions, confusion, or challenges.

Review Process

The research team conducted a multistage review process with each article. Research team coders conducted an initial screen of all manuscripts by reviewing abstracts, ensuring that articles met relevance requirements. In some cases, coders screened the entire manuscript to ensure that inclusion criteria were met. Research team members then coded all articles to capture key characteristics and document details of design, data, sample, analysis, and findings for all studies. During the coding phase, research team members removed articles from the pool if details of the studies indicated the studies were, in fact, not eligible for the topic. If a quantitative study used rigorous methodology, then coders applied additional review standards. Exhibit 1 summarizes the steps of this process.

Exhibit 1.
Literature Review Process

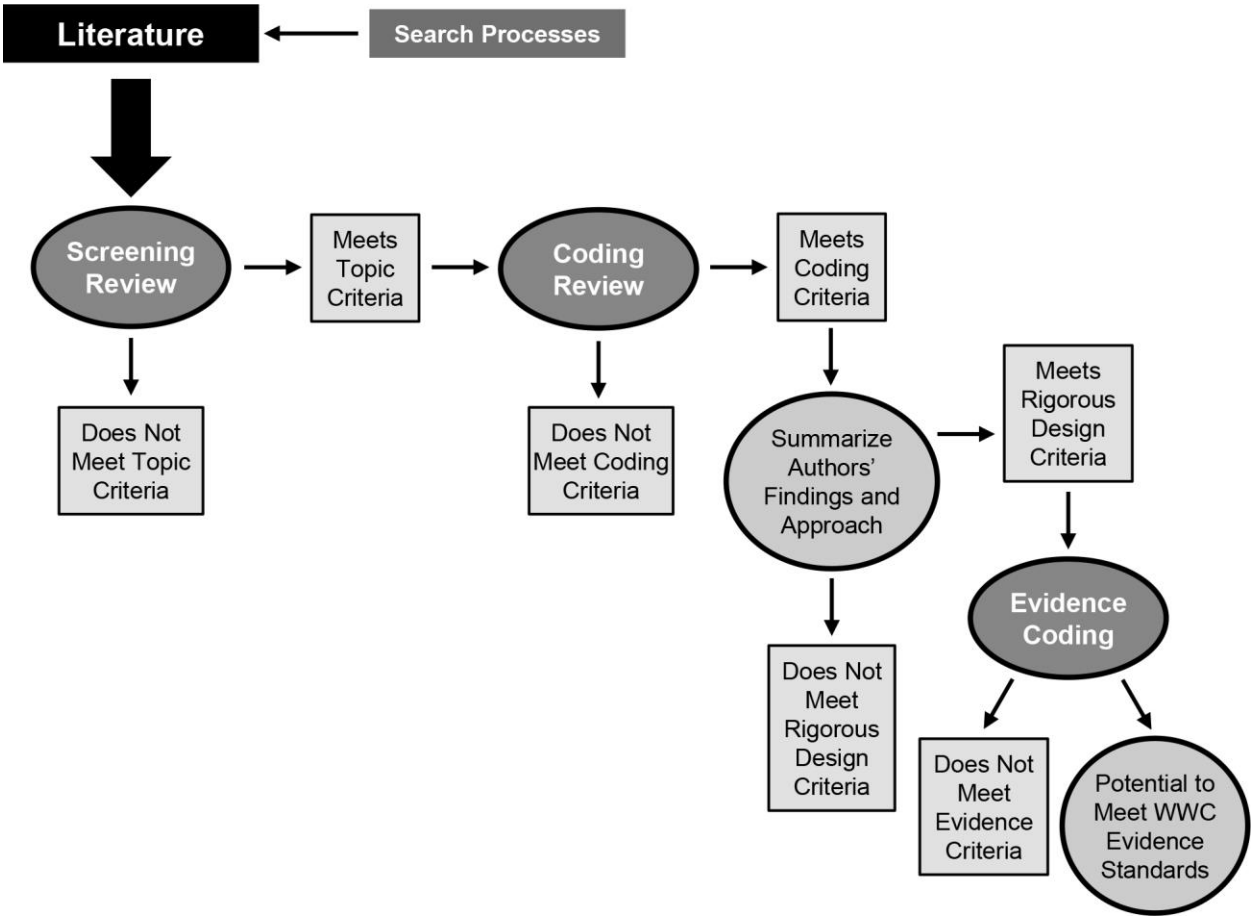

Exhibit reads: Research team coders conducted an initial screen of manuscripts, ensuring that they met relevance requirements. Coders then captured key characteristics and document details of design, data, sample, analysis, and findings for all studies. For studies that used a rigorous design, coders appraised the research methods and data using systematic research standards to determine the level of evidence for the strategy or intervention being studied.

Coding Details

To code content from the policy and theory articles, the research team used NVivo 10, a qualitative software analysis package (QSR International, 2012). Team members drafted a preliminary construct code list, consisting of article elements common across several policy articles. The constructs were defined and coders received training to code article text according to the construct list. For articles with qualitative methodology, the research team documented the aims of the intervention, study methodology, types of data collected, modes of analysis, and findings. Appendix D contains the coding protocols.

There were two pools of studies that used quantitative methods.

- The first pool included rigorous quantitative studies that used at least one comparison group formed using either (1) randomized methods (RCTs) or (2) nonrandomized methods (QEDs). Due to their potential methodological strengths in the use of a comparison group, these studies were reviewed for their evidence of effectiveness.[1]
- The second pool included studies that did not use a comparison group; for example, studies with correlational and single-group pre-test/post-test designs. Because of these studies' designs, they were not reviewed for evidence of effectiveness. Instead, the coding guide was used to capture details about study goals and the author's interpretations of findings.

Results from coding the first pool (rigorous quantitative studies) show that the majority of methodological problems identified with the RCTs and QEDs in this review are related to standards about attrition and baseline equivalence. Attrition refers to the percentage of participants who are missing a post-test measure. Baseline equivalence refers to establishing that, prior to the intervention, participants within the intervention and control conditions in the analytic sample were similar along measurable characteristics (including the outcome measure). Issues with either attrition or baseline equivalence can threaten the strength of a design because it becomes more difficult to confidently attribute the findings to the intervention rather than some other difference between the intervention and control conditions.

[1] Members of the research team, who had previously been certified through the WWC, made use of the *WWC Single Study Review Protocol* (WWC, 2010b) and review standards (consistent with *WWC Procedures and Standards Handbook 3.0;* WWC, 2010a) to determine whether each study has the potential to meet the criteria for being a well-designed study according to the WWC. In this report, we describe studies as having the *potential* to meet WWC group design standards rather than asserting that studies *do* meet standards because the current review is not an official WWC review.

Official reviews conducted by the WWC use author queries to request missing or incomplete information needed to assign a rating or calculate effect sizes. The current literature review did not use author queries because of limited resources. It is possible that more studies would have met evidence standards if author queries had been conducted.

II. P–3 Alignment

Rationale

P–3 alignment aims to coordinate standards, curricula, instructional practices, student assessment, and teacher professional development between the preschool years and the early elementary school years. When implemented as intended, P–3 alignment policies or practices should provide a coherent educational experience as a student progresses from preschool through elementary school (e.g., Halpern, 2013) that could potentially sustain the benefits of preschool (Kauerz & Coffman, 2013). P–3 alignment efforts may include school-based prekindergarten programs and other preschool programs in public or private early care and education settings that partner with the public school system. Because this review includes theoretical literature, the Key Findings section contains additional information about the components and advantages of aligned P–3 models as discussed in the literature.

Literature Search and Screening

At the end of the screening and coding process, 62 articles were reviewed (see Exhibit 2). There were two pairs of articles that contained overlapping content. In the first case, the authors reported the results of one quantitative study in two manuscripts—a working paper (Reynolds, Magnuson, & Ou, 2006) and an article in a published journal (Reynolds, Magnuson, & Ou, 2010). In the second case, a portion of a policy or theory article in a practitioner association resource (National Association of Elementary School Principals, 2011) was reprinted in a different practitioner journal (10 Action Steps, 2011). The final literature review includes 49 policy or theory resources, nine qualitative studies, three quantitative studies, and one mixed-methods study.

Exhibit 2.
Articles Resulting From Literature Search for P–3 Alignment Topic

Literature Search Results	Number of Articles
Total from search	188
Total after screening	66
Total after coding	62[a]
Of 62 articles passing coding stage:	
Policy and theory content coded for themes	49
Studies coded for methods and outcomes	13
Of 13 studies coded for methods and outcomes:	
Qualitative	9
Quantitative—correlational	3
Mixed methods	1

Exhibit reads: The initial total number of articles from the P–3 literature search equaled 188. The number dropped to 66 articles after the screening phase and to 62 articles after the coding phase. Of these, 49 resources contributed unique policy or theory content that the research team coded for themes. An additional 13 resources contained unique studies that the research team coded for methods and outcomes; nine studies were qualitative in nature, three studies were quantitative and used a correlational approach, and one study used a mixed-methods approach.

[a] Studies failed during the coding phase if, for example, the research team discovered that authors discussed the appropriate continuum age and grade range but did not emphasize *alignment* among grades, or if the article was a book review.

There were 49 articles from the literature search that the research team categorized as policy or theory, meaning the article authors did not collect or analyze of data. In general, authors of these articles provided explanations or definitions of P–3 alignment. The authors offered their perspectives of key elements or important characteristics of P–3 practices or programs and reviewed related literature to support their perspectives. Some policy or theory authors included examples of P–3 interventions and approaches, some authors advocated for increasing P–3 approaches (most often through specific policy actions), and other authors provided perspectives on the ways in which P–3 alignment interventions could be facilitated and/or named potential barriers to implementation.

Of the 13 studies that were coded for methods and outcomes, nine used qualitative methods; three used quantitative, correlational methods; and one used mixed methods. Within the qualitative study pool, eight studies used a case-study approach to describe the planning and implementation of P–3 alignment at (1) the state level (Nyhan, 2011; Zellman & Kilburn, 2011), (2) for one or more districts (Jacobson, Jacobson, & Blank, 2012; Marietta, 2010a, 2010b; Marietta & Marietta, 2013a, 2013b), or (3) both state and district levels (Center for the Study of Education Policy, 2012). For these studies, researchers collected implementation data by interviewing stakeholders (e.g., the superintendent, board members, principals, teachers, parents), conducting observations of classrooms, or reviewing extant state or local documents regarding the P–3 approach. One additional qualitative study (Center for Applied Research and Educational Improvement, 2013) provided descriptive data from a cross-section of stakeholders from three districts that participated in a P–3 professional development grant.

Quantitative studies of P–3 alignment are limited, as evidenced by the small number of quantitative studies and the correlational nature of the analyses. One study (Brown & Bogard, 2007) correlated six broad school characteristics[2]—which the authors deemed indicative of a P–3 framework—with students' standardized mathematics and reading achievement, grade retention, and behavior in third grade. Using a similar approach, Reynolds, Magnuson, and Ou (2010) and Reynolds, Magnuson, and Ou (2006) correlated a set of student and school characteristics[3] that they considered part of the P–3 framework, with student outcomes, including reading and mathematics achievement, learning-related behaviors, grade retention, and special education placement. These correlational studies do not provide causal evidence that P–3 approaches improve student outcomes. Furthermore, these broad characteristics and practices only serve as indirect proxy variables for the P–3 approach. The variables in these studies include some characteristics, such as low teacher absenteeism, low teacher turnover, and low student mobility, which are not consistently mentioned in the literature as defining features of a P–3 approach, and do not include other characteristics of the P–3 approach that are defined in the policy and theory literature. Therefore, this review does not discuss the findings of these correlational analyses any further.

---

[2] The six characteristics were (1) principal leadership quality, (2) high academic standards, (3) curriculum planning meetings for teachers, (4) low teacher absenteeism, (5) low teacher turnover, and (6) teacher self-efficacy.

[3] The characteristics included (1) whether children attended preschool before school entry, (2) inclusion of full-day kindergarten; rates of (3) student mobility, (4) highly qualified teachers, (5) parental involvement, (6) amount of reading and language instruction, and (7) average class size.

The mixed-methods study (Bogard, 2006) primarily took a case-study approach to examine P–3 implementation at three schools. This study also conducted analyses to correlate specific school and classroom characteristics or practices at those three schools (e.g., class size, adult-child ratios, specialized teacher training) with classroom quality data.

Findings

Reflecting the state of the research in the field, the findings below focus on theoretical and policy considerations.

Alignment of Standards, Curriculum, Instruction, Assessments, and Environments

Nearly all qualitative studies and policy and theory articles recommend alignment of standards, curriculum, instruction, assessments, and environments across preschool and grades K–3 as an approach for providing high-quality education to students in this grade range. The policy literature calls for both vertical and horizontal alignment of standards, curriculum, and assessment (e.g., Scott-Little & Reid, 2010). Vertical alignment refers to alignment across grade levels, while horizontal alignment refers to alignment within grade.

The literature points to the particular importance of establishing aligned content standards within the P–3 grade range.

Three qualitative studies illustrated specific alignment of content standards in the P–3 grade range. Two of these used a case-study approach to describe P–3 efforts in two districts in New Jersey (Marietta & Marietta, 2013a, 2013b). Using interview data, extant documents, and classroom observations, the authors document that the state developed early learning standards to align with the state's existing content standards for K–12. Researchers highlighted that the state provides lists of approved early childhood curricula and assessments that align with the state P standards (Marietta & Marietta, 2013a). The third study (Center for the Study of Educational Policy, 2012) included a case study of P–3 implementation in the state of Hawaii. To gather information, the study authors conducted in-person interviews with state and local P–3 initiative stakeholders and reviewed secondary data, including documents collected during site visits and through Web searches. Study authors found that Hawaii's efforts involved a school readiness task force that developed preschool standards and later developed broader, but aligned, early learning and development standards that also would align with the Common Core State Standards. Authors in the policy literature explained that many states that adopted the Common Core State Standards have aligned their early learning standards to the Common Core (Guernsey, Bornfreund, McCann, & Williams, 2014).

Curricula and instructional guidance for teachers must be thoughtfully aligned to standards across multiple grades, according to the policy literature.

As examples of this approach, three qualitative studies describe districts that aimed to align curricula across grades. Montgomery County, Maryland, developed its own P–12 curriculum framework and supported alignment through instructional guides for prekindergarten, kindergarten, and later grades with sample lesson plans that align with the district's curriculum framework and state standards (Marietta, 2010a). District administrators in Union City, New

Jersey, worked with teachers to develop a P–12-aligned curriculum (Marietta & Marietta, 2013a). Farrington complex in Oahu, Hawaii, planned to implement a common, published curriculum across the P–3 grade span (Zellman & Kilburn, 2011).

The policy literature points to FirstSchool as a P–3 model that brings together early childhood and elementary education in a single school setting, with alignment of curriculum and instruction (Ritchie, Maxwell, & Clifford, 2007; Ritchie, Maxwell, & Clifford, 2009; New, Palsha, & Ritchie, 2009). The developers of this model note that many children experience discontinuities in curriculum, instruction, classroom setting, and expectations as they move through the P–3 grades, especially during the transition from preschool to kindergarten (New, Palsha, & Ritchie, 2009). For example, although early childhood curricula generally emphasize children's development in a variety of domains, curricula in the later grades place more emphasis on the acquisition of academic content knowledge. According to FirstSchool researchers' observations in a sample of classrooms, children experience a substantial reduction in free-choice time (from 136 minutes to 16 minutes) and an increase in whole-group time (from 76 minutes to 128 minutes) as they transition from prekindergarten to kindergarten (Ritchie, Clifford, Malloy, Cobb, & Crawford, 2010). To facilitate greater alignment, the FirstSchool model employs a curriculum framework to emphasize continuity of student learning goals and professional learning communities for cross-grade instructional planning (New, Palsha, & Ritchie, 2009; Ritchie et al., 2010).

Districts also are implementing common assessment instruments across the P–3 grades.

For example, Montgomery County, Maryland, developed its own diagnostic assessment of reading skills for the K–2 grade range (Marietta, 2010a). Red Bank, New Jersey, selected the Work Sampling System for the P–3 grades (Marietta & Marietta, 2013b). For the Work Sampling System, P–3 teachers assembled portfolios of student work and rated children's performance in the areas of language and literacy, mathematics, and personal and social development, as compared to national expectations and state standards. Teachers shared these portfolios with parents as part of a summary report, which replaced traditional report cards (Marietta & Marietta, 2013b). In the summary report, teachers noted whether the child had made expected progress on the basis of the child's initial performance.

Another concrete approach to alignment is joint professional development and planning time, in which prekindergarten and K–3 teachers come together on a regular basis to focus on curricular and instructional planning.

The policy literature suggests that prekindergarten and K–3 teachers should receive joint teacher preparation and engage collaboratively in planning (e.g., Shore, 2009). Each of the nine qualitative studies and the one mixed-methods study mention joint professional development or planning time; however, the level of detail provided in these case studies varies substantially. Two of the more detailed studies describe Montgomery County, Maryland's approach to P–3 (Marietta, 2010a, 2010b). The district implemented several joint professional development and planning activities. First, early childhood instructional specialists provided teachers with training on standards, curriculum, and assessment. Second, the district developed a 36-hour professional development program for all new P–12 teachers that covered the hallmarks of quality instruction and its importance in helping students reach their full potential. Preschool, Head Start, and kindergarten teachers also participated in supplemental sessions on early learning. As part of

their professional development, teachers conducted classroom observations of their peers. Third, the district developed an online platform for curriculum and lesson planning, which allowed teachers to share lesson planning ideas and link them back to state standards. To allow teacher release time in support of these activities, the district employed a pool of permanent substitute teachers.

Teachers in Union City, New Jersey, met twice per month in cross-grade teams, in addition to meeting twice per week with same-grade teachers to plan instruction and receive mentoring from master teachers (Marietta & Marietta, 2013a). Teachers' participation in these planning meetings allowed time to discuss professional development needs, curriculum implementation, instructional pacing, specific content that proved challenging for students, and effective approaches to teaching that content (Marietta & Marietta, 2013a). An example from the policy literature describes the Birth-to-College initiative, a collaboration between the Urban Education Institute at the University of Chicago and the Ounce of Prevention Fund, in which early childhood educators, elementary school teachers, and family support staff from three schools came together in birth-through-third-grade professional learning communities to foster greater alignment of mathematics, language, and literacy instruction (University of Chicago, Urban Education Institute, & Ounce of Prevention Fund, 2012).

Districts that contract with public and private early childhood education providers to offer preschool often include these providers in district-sponsored professional development to ensure alignment, as described in five of the qualitative studies (Marietta & Marietta, 2013a, 2013b; Marietta, 2010a, 2010b; Zellman & Kilburn, 2011). For example, early childhood education home- and center-based providers may attend the same professional development sessions as district teachers or receive visits from district early childhood education staff or master teachers for training on standards, curricula, and assessment (Marietta & Marietta, 2013a, 2013b; Marietta, 2010b). Such shared professional development may be compulsory or voluntary, and incentives may be provided to encourage participation. For example, a district case study describing a P–3 initiative in Bremerton School District in the state of Washington (Marietta, 2010b) described a "district-endorsement" for early childhood education providers who attended district-sponsored professional development sessions. Providers, in turn, can use this district endorsement to market their early childhood education programs.

The literature suggests that, to support P–3 alignment, classroom environments should be similar: All classes should be small; preschool and kindergarten, in particular, should have similar classroom structures and environments.

Two qualitative studies provided specific case-study examples. Montgomery County, Maryland, reduced K–2 class sizes to 15 in high-need schools, as part of P–3 reforms (Marietta, 2010a). Union City, New Jersey, directed kindergarten teachers to arrange their classrooms into learning centers, which are similar to those found in preschool classrooms, rather than in rows of desks (Marietta & Marietta, 2013a). The theory and policy articles also advocated for small classes with similar structures (e.g., Grantmakers for Education, 2006; Black, 2008; Bogard & Takanishi, 2005; Committee for Economic Development, 2012; Howard, 2008; Rice, 2008a; Rice, 2010). For example, Reynolds and Ou (2006) described the Chicago Child-Parent Centers (CPC) program, which attempted to create greater continuity in classroom environments for children participating in the program. During the preschool year, both a teacher and an aide

staffed classrooms, with a maximum of 17 students. During the K–3 period, participating students continued to experience small class sizes, with a maximum of 25 students and two staff. The class sizes offered through CPC were considerably smaller than typical first- through third-grade classrooms in Chicago, which enrolled 35–40 students with just one teacher.

Kindergarten readiness standards and kindergarten entry assessments can serve as mechanisms to facilitate alignment from preschool to kindergarten.

The policy literature suggests kindergarten readiness standards and associated kindergarten entry assessments as a model strategy for alignment between early education and elementary education (Tout, Halle, Daily, Albertson-Junkans, & Moodie, 2013). Kindergarten readiness standards provide early care and education providers with further guidance regarding the expectations young children will encounter at school entry, and kindergarten entry assessments provide kindergarten teachers with diagnostic data on individual students that they can use to plan instruction (Center for the Study of Educational Policy, 2012; Tout et al., 2013; Zellman & Kilburn, 2011). Two qualitative case studies mention the role of kindergarten entry assessments in alignment in the context of Hawaii's State School Readiness Assessment (Center for the Study of Educational Policy, 2012; Zellman & Kilburn, 2011). Kindergarten teachers use one assessment to look at overall readiness of children at the classroom level and another assessment to measure the readiness of individual students. Aggregated information is shared publicly to improve the education of young children (Center for the Study of Educational Policy, 2012). The Center for the Study of Educational Policy (2012) describes how Pennsylvania planned to house kindergarten readiness assessment data in the state's longitudinal K–12 student data system, in addition to integrating the state's early childhood data system for children ages zero to five with the K–12 system.

According to the theory and policy literature, the ultimate goal of alignment is to ease children's transitions into school and across grade levels.

Examples of specific transition practices include (1) the transfer of records from prekindergarten to kindergarten, (2) kindergarten classroom visits for children, or (3) parent orientations prior to the beginning of school (Kagan et al., 2006; Tout et al., 2013). Children's entrance into elementary school is an important transition in early childhood that can set the stage for future success or failure (Demanchick, Peabody, & Johnson, 2009; Human Capital Research Collaborative, 2014a; New, Palsha, & Ritchie, 2009; Tout et al., 2013). Numerous theory and policy articles emphasize the importance of parental involvement and communication between teachers and parents in the transition process (ABCs of Early Education, 2013; Goldstein & Bauml, 2012; Groark, Mehaffie, McCall, & Greenberg, 2007; New, Palsha, & Ritchie, 2009; Rice, 2008b; Tout et al., 2013).

Authors point to the Chicago CPC program as an example of a P–3 intervention program that includes formal transition practices (Human Capital Research Collaborative, 2014a, 2014b; Reynolds & Ou, 2006). The CPC program offered early childhood education and family support services to low-income families, and follow-up services through third grade in order to sustain the effects of the preschool intervention (Human Capital Research Collaborative, 2014a, 2014b). CPC programs were purposely based in public schools with the aim that participating students would experience easier transitions as they moved from preschool to kindergarten **(**Human

Capital Research Collaborative, 2014b). Specific transition practices in the CPC program included maintaining the same staff leadership team as children age through the program, supporting communication between CPC head teachers and school principals, developing a continuity plan, and planning for cross-grade activities **(**Human Capital Research Collaborative, 2014a).

Teacher Education and Qualifications

Numerous policy articles call for establishing similar teacher education and training requirements across preschool and elementary education job positions, and several qualitative studies provide examples of this practice. Authors suggest that preschool teachers should earn bachelor's degrees, hold certification, and receive compensation that is equivalent with that of elementary teachers. Furthermore, they suggest that K–3 elementary school teachers should receive more training in early childhood development.

Some authors of policy and theory articles recommend that preschool teachers should earn the same educational credential as elementary teachers, namely a bachelor's degree. Authors also argue for equal compensation for preschool teachers and elementary school teachers.

Three qualitative case studies describe P–3 efforts in which preschool teachers held bachelor's degrees and had salary parity with their peer teachers in the K–3 grades (Marietta, 2010a; Marietta & Marietta, 2013a, 2013b). Two case studies document this approach in New Jersey, where preschool teachers in Union City and Red Bank must hold a bachelor's degree and P–3 certification and receive the same pay as other elementary school teachers (Marietta & Marietta, 2013a, 2013b). Both of the New Jersey districts partnered with private and nonprofit early childhood education programs to deliver preschool, and the teachers in these out-of-district programs met the same education requirements and received the same pay as teachers inside the district. Thus, the approach in these districts maintained consistent standards across settings. Another case study describes a Montgomery County, Maryland, preschool program that was part of a P–3 strategy to increase student achievement in later grades. The district hired only certified teachers with a bachelor's degree, employing them as regular teachers who earned the same salary as other district teachers (Marietta, 2010a).

The policy literature further recommends that elementary school teachers receive training in early childhood development (Rice, 2008a; Kauerz, 2006; Takanishi & Kauerz, 2008), although the qualitative studies do not provide any examples of this approach.

The creation of P–3 teacher certification programs provides an opportunity to build a shared educational philosophy among early childhood educators and elementary school teachers of the K–3 grades, thus increasing alignment.

Two case studies and two policy articles document the development of P–3 teacher certification programs in New Jersey, as mandated by a Supreme Court of New Jersey ruling in an education equity case (Rice, 2007; Marietta & Marietta, 2013a, 2013b; Mead, 2009). Graduates of these training programs possess a bachelor's degree with a P–3 endorsement. The court ruling required P–3 certification only for prekindergarten teachers, but the Advocates for Children of New

Jersey recommended that all new K–3 teachers also be required to hold the P–3 certification in order to address the issue of alignment between prekindergarten and K–3 (Rice, 2007). Drawing on interviews, focus groups, and document review, two other qualitative studies highlight the work of Hawaii's P–20 Partnership for Education, a working group that brings together representatives from early childhood, K–12, and higher education (Center for the Study of Education Policy, 2012; Zellman & Kilburn, 2011). The P–20 Partnership worked with community college and university faculty to increase course offerings in early childhood education, and established a P–3 graduate certificate program. Teachers at P–3 pilot sites in the state were encouraged to enroll in the certificate program and received full tuition scholarships from the P–20 Partnership. The certificate program included coursework credit hours that could later count toward a P–3 master's degree, if teachers chose this pathway.

Data-Driven Instructional Planning

Numerous policy articles recommend the creation of systems that link individual student data from public and private early childhood programs, particularly preschool programs, to students' public school data so that elementary teachers have more complete and accessible information about students' learning trajectories. With access to these data, professional development on their use, and cross-grade planning time, P–3 educators could better tailor instruction to meet students' needs.

The theory and policy literature recommends development of longitudinal P–12 or P–20 data systems that link data from public and private early care and education programs to public school data.

Longitudinal data systems would allow administrators and teachers to have more complete and accessible information about students' learning trajectories than current approaches to collecting and storing student data (10 Action Steps, 2011; Hernandez, 2012; Kauerz & Coffman, 2013; NALEO Education Leadership Initiative, 2008; Lesaux, 2010; The Pre-K Coalition, 2011a; Rice, 2010). One author calls on the federal government to convene a national advisory group to create guidelines for the development of state longitudinal data systems, and state governments to establish new laws and regulations that allow for data sharing while protecting student confidentiality (Hernandez, 2012).

The policy literature further suggests that districts may use longitudinal data systems to inform teacher performance evaluation (Buenafe, 2011; Guernsey et al., 2014; Kauerz, 2009; Takanishi & Bogard, 2007; Takanishi & Kauerz, 2008). However, Guernsey and colleagues (2014) suggest that caution is warranted because many early childhood assessments are formative or diagnostic in nature and are not validated for use in teacher evaluation. Similarly, some observation tools used to evaluate teachers have not been validated for early childhood settings (Guernsey et al., 2014). Thus, several states are field testing observation tools (Guernsey et al., 2014). To address concerns about prekindergarten teacher performance and student outcomes, some states and localities are also developing or refining their quality rating and improvement systems, which rate the quality of early learning programs on the basis of teacher qualifications; teacher-child ratios; class size; and, in some cases, measures of teacher-child interactions (Buenafe, 2011; Guernsey et al., 2014).

The literature emphasizes the role of student data in P–3 instructional planning and professional development.

To support data-driven instructional planning in P–3, the theory and policy literature calls on administrators and principals to provide school-wide assessment data, as well as disaggregated data by student subgroups (defined by demographic group, classroom, and grade level) (Kauerz & Coffman, 2013). These data would allow teachers to monitor student progress and address achievement gaps (Kauerz & Coffman, 2013). For teachers to make efficient use of assessment data for curricular and instructional planning, policy authors suggest that teachers need professional development on the assessment instruments, as well as any data systems where assessment data are stored, and regular cross-grade planning time with other teachers (10 Action Steps, 2011; ABCs of Early Education, 2013; Kauerz & Coffman, 2013; Mead, 2009; National Association of Elementary School Principals, 2011; Lesaux, 2010).

Two qualitative studies document such systems. A descriptive study of a P–3 professional development initiative in Minnesota illustrates the role of student data in professional development and planning (Center for Applied Research and Educational Improvement, 2013). The Urban Education Institute at the University of Chicago designed and delivered Minnesota's professional development initiative to improve early literacy instruction. As part of the program, coaches taught teachers to administer assessments and use assessment data to plan instruction. Coaches and teachers had access to individual students' scores on specific subdomains of early literacy related to oral language and familiarity with print. Based on the assessment data, coaches taught P–3 professional development workshops on specific instructional strategies and recommended texts for guided reading groups. For this study, the researchers conducted 54 interviews with districts and school administrators, teachers, and literacy coaches. Participants reported that the initiative led to improved communication among teachers of different grades and improved student performance.

A second qualitative study, drawing on interviews and document review, recounts early childhood teachers' efforts at two P–3 pilot sites in Hawaii to assemble student data in the form of student portfolios, with information on children's families and samples of their work to document learning and development (Center for the Study of Education Policy, 2012). These student portfolios were shared with kindergarten teachers to inform instructional planning and ease children's transitions into elementary school.

Administrative Leadership

Several policy articles and qualitative studies suggest that school district administrators can support the implementation of P−3 initiatives through the management practices they put in place. Specific leadership considerations include the following: (1) involving early childhood education providers and K–3 teachers in planning P–3 initiatives, (2) implementing the planned elements of P–3 initiatives with fidelity, (3) specifying measurable student achievement benchmarks, and (4) holding principals and teachers accountable for achieving benchmarks. Two study authors also link similar principal management practices to implementation of P–3 initiatives.

District administrators involve early childhood education providers and K–12 teachers in the planning of P–3 initiatives to obtain input and encourage buy-in for the initiative by both sets of educators.

The policy literature stresses administrators' roles in building cross-sector collaboration and fostering teacher involvement to implement P–3 efforts (10 Action Steps, 2011; Kauerz, 2009; Kauerz & Coffman, 2013; National Association of Elementary School Principals, 2011). One qualitative study highlighted that the superintendent of Red Bank, New Jersey, worked with a committee of teachers to develop a strategic plan for the district's early grades (Marietta & Marietta, 2013b). When teachers and administrators expressed reservations about overhauling the district's approach to curriculum, instruction, and assessment in the P–3 grades, the superintendent arranged a site visit and several meetings with another district that had already adopted a similar approach. As a result of these meetings, the majority of teachers and administrators agreed the changes would be beneficial. Another case study describes an experience in Union City, New Jersey, where the district administrator gave teachers the authority to write the district's P–12 curriculum and align it across grades. Teachers update the curriculum annually during a summer planning process, which includes cross-grade meetings (Marietta & Marietta, 2013a).

District administrators maintain high standards for P–3 initiatives by holding principals and teachers accountable for implementing the planned elements of the P–3 initiative.

In Union City, New Jersey, and Montgomery County, Maryland, administrators from the central office conducted regular classroom visits to P–3 classrooms to observe instructional practices and ensure that teachers were implementing the planned curriculum (Marietta & Marietta, 2013a; Marietta, 2010a). In Montgomery County, prekindergarten and kindergarten teachers were expected to make their instructional plans and summaries of student performance data available for principal review during classroom observations (Marietta, 2010). Administrative guidelines in Union City directed principals to conduct daily walk-throughs to guide instructional planning and future professional development (Marietta & Marietta, 2013a). In addition, when Union City first adopted the P–3 approach, master teachers conducted walk-throughs to check that teachers had implemented the district's plan to arrange kindergarten classrooms into learning centers that are similar to those of a preschool classroom (Marietta & Marietta, 2013a).

District administrators set high expectations for P–3 initiatives when they establish specific student achievement benchmarks and gather data to measure progress toward the benchmarks.

The policy literature suggests that student achievement benchmarks are needed in order to assess the results of P–3 initiatives (Guernsey et al., 2014; Kauerz, 2009). Because the results of early education are not assessed in the same manner as the later elementary grades and beyond, district administrators must play a leadership role in setting student achievement benchmarks for the P–3 grade range (Kauerz, 2009; The Pre-K Coalition, 2011b). The establishment of student achievement benchmarks also helps principals focus on the P–3 grades rather than focusing more exclusively on the later grades where standardized testing occurs (Guernsey et al., 2014). One qualitative study described Montgomery County, Maryland, where the superintendent sought to ensure that students were reading proficiently by third grade and 80 percent of high school

students met college readiness benchmarks (Marietta, 2010a). The district implemented a professional development system for all P–12 teachers, regular formative assessments to track student progress, and teacher accountability measures. Students were assessed using Maryland's kindergarten readiness assessment, a district-created early literacy assessment, and multiple measures for mathematics to inform instructional planning and track student progress (Marietta, 2010a, 2010b). The district established an Office of School Performance, which administered a peer assistance review program for the district's P–12 teachers. Through this program, consulting teachers advised new teachers and struggling veteran teachers on classroom practice. At the end of the year, consulting teachers made employment recommendations to an oversight panel governed by district and union representatives.

The literature emphasizes the importance of principal leadership in implementation of P–3 initiatives.

For example, a case study of Union City, New Jersey, documents district administrators' expectation that principals will implement and monitor components of the P–3 initiative (Marietta & Marietta, 2013a). The district central office provides principals with guidance describing their responsibility for distributing assessment data to teachers for cross-grade instructional planning, conducting daily classroom visits to observe instruction, and organizing regular cross-grade teacher planning meetings (Marietta & Marietta, 2013a). District administrators conduct school-wide assessment team visits to observe all classrooms within schools and hold in-person meetings with principals to discuss the results of classroom observations and assessments. Principals, in turn, develop specific plans to improve instruction in areas where student learning is weak, typically using additional teacher supports, such as master teachers. In addition to placing emphasis on the importance of principal leadership (Black, 2008; Bogard, 2006; Bogard & Takanishi, 2005; Howard, 2008; Brown & Bogard, 2007; Takanishi & Kauerz, 2008), the policy literature also suggests that training in early childhood education is important preparation that equips administrators and principals to lead the development of a coordinated P–3 system within their building or district (Advocates for Children of New Jersey, 2010; Donovan, 2010; Guernsey et al., 2014; NALEO Education Leadership Initiative, 2008; Rice, 2007).

Challenges

According to the policy literature, the following challenges must be addressed if P–3 initiatives are to be more widely implemented: (1) policies that inhibit the blending of federal, state, and local sources of funding to support P−3 initiatives; (2) instability of preschool funding; (3) resistance by practitioners to integration of preschool and the K–3 grades; and (4) the organization of elementary education classrooms, buildings, and enrollment.

The lack of a unified and stable funding stream is a barrier to the creation of sustainable, unified P–3 systems.

Four of the qualitative studies document challenges to blending funding at the district level (Marietta & Marietta, 2013a, Jacobson et al., 2012; Marietta, 2010a; Nyhan, 2011). One case study describes Montgomery County, Maryland, which funded preschool using Head Start, Title I, and other local funds set aside through collaboration with the Montgomery County Department

of Health and Human Services, County Council, and County Executive (Marietta, 2010a). The funding sources supporting specific preschool slots varied with children's family income levels. Head Start funding supported slots for the lowest income children. Children at the higher end of the income spectrum were able to participate because their teachers were high school students in an early education internship program.

Seattle's preschool program relied on funding and technical assistance from three different philanthropic foundations (Nyhan, 2011). According to the author, goals across organizations sometimes differed. For example, one foundation objected when the district housed a program for children with social-emotional needs in a P–3 school because the foundation wanted to build a model school. Seattle's experience demonstrates that it can be difficult to fulfill the requirements and desires of multiple funders.

The theory and policy literature further describes the funding challenges facing P–3 initiatives. Separate federal funding streams for preschool and elementary school have prevented easy utilization and combination of funds for P–3 efforts (e.g., Advocates for Children of New Jersey, 2010; Gates Foundation, 2011; Jacobson, 2009). In addition, states have varying policies and practices regarding the funding and availability of preschool (Halpern, 2013; NALEO Education Leadership Initiative, 2008; Takanishi & Bogard, 2007; Takanishi & Kauerz, 2008), and some states and districts have turned to funding preschool through private monies or efforts, such as tax levies (Garland, 2011; Maeroff, 2003; Mead, 2009; NALEO Education Leadership Initiative, 2008). These funding streams have varying standards and regulations, which complicate efforts to unite preschool and elementary school (Advocates for Children of New Jersey, 2010; Jacobson, 2009; Rice, 2007; Kagan & Kauerz, 2010; Maeroff, 2003; Kauerz & Coffman, 2013; NALEO Education Leadership Initiative, 2008; National Association of Elementary School Principals, 2011). To remedy these barriers, the policy literature calls on government to enable more seamless coordination and blending of federal and state funding streams for early childhood education services (10 Action Steps, 2011; King, 2006; The Pre-K Coalition, 2011a; National Association of Elementary School Principals, 2011).

One qualitative study describes two case studies that illustrate the impact of unstable preschool funding on school districts that attempted to operate P–3 programs with discretionary funding (Jacobson et al., 2012). The school district in Evansville, Indiana, had operated a preschool program for 13 years with Even Start funding. When Congress cut funding for Even Start in 2011, the district could no longer maintain the preschool program and the children served were forced to enroll in other Head Start and early education programs outside of the school system and the P–3 initiative. Another district in Cincinnati, Ohio, had relied on Ohio's Early Learning Initiative—a state funding stream for early education supported by Temporary Assistance for Needy Families—to fund preschool but found that it could no longer effectively operate the program when the state made substantial funding cuts.

K–3 administrators, teachers, and early childhood providers may resist the idea of combining or aligning preschool with grades K–3 because there is a perception of significant philosophical differences between early childhood and elementary grade teachers.

As noted earlier, while early childhood curricula generally emphasize children's development in a variety of domains, some stakeholders believe that curricula in the later grades only emphasize the acquisition of academic content knowledge (New, Palsha, & Ritchie, 2009). Takanishi (2010) asserts that prekindergarten educators must shift their perspective to be more inclusive of a focus on early academic skills and avoid portraying K–3 education as a "skill-and-drill" experience that focuses only on content knowledge. K–3 educators can adopt a whole-child philosophy similar to early childhood teachers, and early childhood teachers can include developmentally appropriate coverage of content such as mathematics and science (Jacobson et al., 2012; Takanishi, 2010).

The organization of elementary education classrooms, buildings, and enrollment also can be a challenge to creating P–3 models.

One qualitative study describes an example in which kindergarten teachers, who objected to Union City, New Jersey's decision to rearrange their classrooms into learning centers, involved the local teachers union in their dispute (Marietta & Marietta, 2013a). A second study highlights two districts in Hawaii—Nanakuli-Wai'anae and Farrington—that arranged for preschool teachers to share portfolios of children's work with their future kindergarten teachers. The districts learned that some kindergarten teachers had not received the portfolios because principals and other school staff did not know the purpose of the portfolios or who was to receive them (Center for the Study of Educational Policy, 2012). Finally, a third qualitative study describes difficulty in building connections across preschool and K–3 in a school that hosted Head Start programs because children left to attend kindergarten in other elementary schools (Jacobson et al., 2012).

Conclusion

When implemented as intended, P–3 alignment policy or practices should provide a coherent educational experience as a student progresses from preschool through elementary school (e.g., Halpern, 2013). This could potentially sustain the benefits of preschool (Kauerz & Coffman, 2013). Extant literature, including 49 policy and theory articles, nine qualitative studies, two quantitative studies, and one mixed-methods study, recommends alignment of standards, curriculum, instruction, assessments, and environments across preschool and grades K–3. Authors suggest that establishing similar teacher education and training requirements, and equivalent compensation across preschool and elementary education job positions, would support P–3 alignment. The literature also indicates that creating longitudinal student data systems that integrate prekindergarten with K–12 data, providing P–3 teacher professional development on data use, and offering cross-grade planning time would support the use of student assessment data in P–3 instructional planning. In addition, district administrators and principals can support the implementation of P−3 initiatives by involving teachers in the planning process, ensuring fidelity of implementation, measuring student achievement benchmarks, and holding administrators and teaching staff accountable. Within the literature, some authors point out challenges to P–3 alignment implementation; these include policies that inhibit the blending of funds, instability of preschool funding, and resistance among practitioners to integration of preschool and the K–3 grades.

III. Differentiated Instruction

Rationale

Differentiated instruction is a way to meet students' diverse needs (Parsons, Dodman, & Burrowbridge, 2013) by having teachers deliver instruction through multiple modes or at multiple levels (Lawrence-Brown, 2004). As Tomlinson and colleagues (2003) explain, differentiated instruction, or "academically responsive instruction," aims to ensure that all students in a classroom have equal access to quality instruction, despite their varying levels of skills, motivation, interests, or their heterogeneous economic, cultural, and linguistic backgrounds. Differentiation requires that teachers carefully plan instruction to account for the variation of learners in their class (Tomlinson, 1999) and make adaptations to meet student needs (Parsons, 2012; Parsons et al., 2013). In a differentiated instruction delivery model, there are various ways to be responsive to the needs of individuals or groups of students—sometimes referred to as individualization of *content*, *process*, or *product* of instruction (Anderson, 2007; Parsons et al., 2013; Stanford & Reeves, 2009; Tomlinson et al., 2003). For example, teachers could use varying instructional practices or strategies with students, change the content to be more complex or simplified for particular students, adapt or modify curricular resources or materials, or change the procedures for student evaluation (e.g., Brimijoin, 2005; Tomlinson et al., 2003).

One explanation for why preschool effects diminish in early elementary school is that children who make early gains in preschool may not have the opportunity to maintain their rate of learning because early elementary instruction is oriented to students with the lowest level skills and therefore does not capitalize on the skills that some students have upon school entry (Kauerz, 2006). As students make the transition to elementary school, it appears to be important that the content and instruction they encounter is challenging enough. Using the Early Childhood Longitudinal Study–Kindergarten Cohort (ECLS-K) teacher survey and child achievement data, Claessens et al. (2013) analyzed the relationship between content coverage and end-of-kindergarten reading and mathematics achievement. The study was not focused on differentiation, but researchers found that when kindergarteners, whether they attended preschool or not, have more exposure to advanced content and less exposure to basic content, there are larger achievement gains. However, exposure to basic content is much more frequent.

In practice, differentiation and providing more challenging instruction for some students may be difficult for teachers. Survey and observational data have shown that teachers generally make few adjustments to instructional and curricular practice to address the needs of advanced learners in a regular classroom (e.g., Archambault et al., 1993; Westberg, Archambault, Dobyns, & Salvin, 1993).

Literature Search and Screening

After the screening and coding process, the review included 21 studies: 17 quantitative studies and four qualitative studies (see Exhibit 3).

Exhibit 3.
Literature Search Results for Differentiated Instruction Topic

Literature Search Results	Number of Articles
Total from database search	506
Total after initial screening	71
Total after full text screen, removing studies that only focus on a low-achievement group	68
Total after coding	21[a]
Of 21 studies coded for methods and outcomes:	
Quantitative:	
Descriptive	1
Quasi-experimental	6
Randomized controlled trial	7
Single-group pre-test/post-test	3
Qualitative	4
TOTAL	21

Exhibit reads: The initial total number of articles from the differentiated instruction literature search equaled 506. The number dropped to 71 articles after an initial screening phase and to 68 articles after another screening that removed studies that only focused on a low-achievement group. There were 21 studies that passed the coding phase. Of these, four studies were qualitative in nature. There were 17 quantitative studies. Of these, one used a descriptive approach, six used a quasi-experimental approach, seven used a randomized controlled trial, and three were a single-group pre-test/post-test design.

[a] Studies failed during the coding phase if, for example, the research team discovered during a more detailed reading that the study did not use an approach consistent with our definition of differentiated instruction or if the study sample did not meet criteria.

Thirteen studies used RCT or QED designs and were eligible for the full evidence of effectiveness review. Four quantitative studies used other designs that were not eligible for the full review, although coders still captured descriptive information about these studies using the additional characteristics and structured abstract sections of the coding guide. These four studies are included in the following findings, but with less confidence in the attribution of effects to the intervention. Further information on the quantitative analyses for the 13 studies eligible for evidence of effectiveness review can be found in Appendix E.

Four studies used qualitative designs and were not eligible for the full review. Reviewers captured research design and findings information on these studies using the additional characteristics and structured abstract sections of the coding guide. These studies focused on processes and strategies for implementing differentiated instruction for mathematics instruction and on researcher perceptions of factors that facilitate or hinder implementation. These studies are described in this report to provide additional insight into differentiated instruction implementation.

Findings

A total of 21 studies met screening criteria for inclusion in this review, including 17 quantitative studies and four qualitative studies. Of the 17 quantitative studies, seven used RCTs, six used QEDs , and four used other non-rigorous designs (i.e., descriptive and single-group pre-test/post-test designs) to examine the effects of differentiated instruction on achievement for students in kindergarten or grade 1. However, most of these studies have methodological issues that diminish the level of confidence in the study to demonstrate causal evidence of effectiveness.

The studies included in this review examined a variety of content areas. Most of the studies (14) that met screening criteria for topic relevance focused on reading instruction (seven RCTs, five QEDs, and two quantitative studies with other designs). Three studies (one QED and two single-group pre-test/post-test studies) evaluated the impact of differentiated instruction on writing outcomes. Four qualitative studies examined implementation of differentiated instruction in mathematics.

In the studies reviewed for this report, differentiated instruction practices and programs were offered as individualized or group instruction. In *individual, child-level differentiation*, teachers differentiated instruction based on an individual student's specific needs. Instruction may have occurred individually, in a small-group setting, or in a whole-classroom setting, but the lesson planning aimed to address individual student needs rather than the needs of a group. In *differentiation for groups* of children, researchers divided students into small groups of children who were similar along a specific dimension and differentiated instruction was based on the perceived overall needs of the group.

Individualized Differentiated Instruction on Reading

The intervention package, Individualized Student Instruction with Assessment to Instruction, demonstrated positive effects on reading outcomes in six RCTs. One substudy in one of the RCTs has potential to meet WWC research standards for strong casual evidence.

The intervention package examined by these six studies contains two main components—Individualized Student Instruction (ISI) and Assessment to Instruction (A2i). These tools provide training and professional development to teachers on how to individualize literacy instruction in the classroom using the recommendations and planning strategies provided by A2i Web-based software. ISI and A2i aim to improve a teacher's ability to differentiate reading instruction based on individual students' needs. The A2i software uses students' literacy outcome scores on the Woodcock-Johnson III Tests of Achievement (Letter-Word Identification and Picture Vocabulary subtests) to develop strategies that teachers then use to differentiate instruction in the classroom. The A2i software also uses the scores to divide students into smaller groups based on their skills and needs. In this way, small-group instruction also can be used in the classroom based on ongoing student achievement information. A description of the body of research on the ISI and A2i bundled intervention follows.

Connor and colleagues have produced five reports on studies that used an RCT design. Based on a review of the published articles, the authors have produced four of these five reports based on

the same RCT sample of 10 Florida schools in grades 1–3 (Al Otaiba et al., 2011; Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007; Connor et al., 2009; Connor et al., 2010) and one report based on an independent randomization of a different sample of teachers in grades 1–3 in north Florida (Connor et al., 2013). All five studies have methodological issues that are discussed later in this section. Hence, most of the findings described below should be interpreted with caution.

The goal of Connor, Morrison, Fishman, Schatschneider, and Underwood (2007) was to assess whether Individualized Reading Instruction (using A2i)[4] had an effect on students' reading achievement relative to other types of small-group reading instruction. To answer this question, the researchers randomly assigned schools to either an intervention condition or a control condition. All teachers were expected to dedicate time for a daily 90-minute reading block. In the intervention condition, teachers received training on planning and implementing Individualized Reading Instruction using A2i. In the control condition, teachers were expected to use small groups as suggested by school policies. The outcome measure used was a test of students' language and literacy skills, the Woodcock-Johnson III Tests of Achievement. The findings showed that the intervention group achieved stronger reading growth relative to students in the control group. Students in the intervention group exhibited reading growth that was an average of 2.63 points higher than the reading growth for students in the control group (see Exhibit E1). The authors did not report effect sizes, standard deviations, or sample sizes by group.[5]

Connor and colleagues (2009) investigated the implementation of the ISI intervention to determine if teachers who received the intervention individualized instruction closer to the A2i recommendations than comparison group teachers. The study also investigated whether intervention students had greater reading growth than comparison students across different levels of precision between observed instruction and A2i-recommended instruction. In this study, 10 schools were randomly assigned to either the intervention condition, where schools received training on how to individualize literacy instruction using A2i, or the control condition, where schools were put on a waitlist to receive the training the following year. The district required all of the schools in the study to provide a two-hour language instruction block, with 45 minutes devoted to small-group instruction. Authors measured reading outcomes using the Woodcock-Johnson's Letter Word Identification, Passage Comprehension, and Picture Vocabulary subtests. The authors reported that intervention teachers individualized instruction closer to the A2i recommendations than comparison teachers. The authors also reported that when students spent more time engaging in teacher/child-managed, meaning-focused instruction, both their passage comprehension skill growth and their letter word reading growth were greater. However, the authors did not find statistically significant effects on reading outcomes when testing the interaction between the treatment condition and the precision with which the observed instruction matched the A2i-recommended instruction (see Exhibit E2).

[4] Between 2007 and 2009, the researchers changed from using the term Individualized Reading Instruction to the term Individualized Student Instruction (ISI). The intervention appears to be the same but with a different label.

[5] The review of this study only relied on information reported in the published article.

Connor and colleagues (2010) continued their investigation of the ISI intervention with the goal of investigating whether ISI use in classrooms contributed to growth in student self-regulation, as measured by the Head Toes Knees Shoulders (HTKS) task. The findings showed no main effect of the intervention on self-regulation growth; there was no significant effect on their HTKS score gains from fall to spring, controlling for initial fall literacy scores. However, the authors did find that the average difference in self-regulation between the intervention and comparison groups increased as the classroom teachers' use of A2i increased. In other words, there was an interaction between amount of A2i use and student self-regulation outcomes. Overall, the authors concluded that self-regulation may be malleable during the early years of school and that focusing on the classroom environment in ways that increase self-regulation may be helpful for student success and academic achievement (see Exhibits E3 and E4).

Connor and colleagues conducted another school-level RCT on ISI and A2i in 2011. The goal of this study was to determine if there were interactions between child characteristics and instruction type that caused outcome variation. The study asked two student-level research questions. The first question evaluated the main effect of individualizing literacy instruction using A2i recommendations compared to "business-as-usual" literacy instruction. The second question investigated the difference in impact for children with different background characteristics. In particular, the authors measured differences based on reading skills, school socioeconomic status, and special education status. The measure used for language and literacy skills was the Woodcock-Johnson III Tests of Achievement. Intervention students demonstrated greater and statistically significant gains in Letter Word Identification subtest scores than did students in the control condition. The authors also found that the intervention may be *less* effective for students receiving special education. Finally, the authors found that there was a greater impact on scores for students with lower pre-intervention scores (students at the 25th percentile) than for students with higher pre-intervention scores (students at the 75th percentile; (see Exhibit E5).

Al Otaiba and colleagues (2011) investigated the ISI intervention in a kindergarten sample (ISI-K). The authors aimed to determine the effect of ISI and A2i on kindergarten students' reading scores. For the kindergarten outcomes, the authors measured reading scores using the following measures: Woodcock-Johnson III Letter Word Identification, Woodcock-Johnson III Word Attack, AIMSweb's Letter Sound Fluency, Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Phoneme Segmenting Fluency (to measure phonemic awareness), and DIBELS Nonsense Word Fluency (to measure phonics and decoding). The intervention and comparison classrooms had a common professional development program from the Florida Progress Monitoring and Reporting Network, which included a daylong workshop on response to intervention and individualized instruction, training on material and games, and interpreting student data. The intervention group also received training and ongoing professional development on using the A2i software. The authors reported a large overall positive effect on literacy outcomes (see Exhibits E6 through E9) and stated that individualizing instruction can lead to stronger student literacy outcomes at the end of kindergarten within a diverse group of students.

Because Connor et al. (2007, 2009, 2010, and 2011) and Al Otaiba et al. (2011) used RCT designs for these studies, the literature review research team reviewed them for evidence of effectiveness, which revealed some methodological issues. In all five studies, the authors did not

report the level of attrition (i.e., the percentage of students in the sample who were missing a value for the outcome). If this percentage was high, it may put into question the similarity between groups created by random assignment prior to the delivery of differentiated instruction. In Connor et al. (2007 and 2009) and Al Otaiba et al. (2011), the authors did not report enough baseline data (sample sizes, means, and standard deviations) to show whether the two groups were equivalent prior to the intervention. The analytic sample for the Connor et al. 2010 study was found to be *unequal* on academic measures prior to the intervention. These differences between the groups prior to the delivery of differentiated instruction may serve as an alternative explanation to differentiated instruction for the differences between groups on the outcome. Finally, the authors did not report sample sizes for the fall and spring assessment data in the Connor et al. 2011 study, making it impossible to determine whether the baseline data presented in the study represent the analytic sample. For these reasons, the reported estimates for these five studies should be interpreted with caution.

In 2013, Connor and colleagues examined the ISI and A2i intervention once again, with four purposes: (1) to determine whether previous single-grade studies and the algorithms used by the A2i software to make differentiated instruction recommendations could be replicated; (2) to investigate if the effect of ISI accumulates as students receive more years of the intervention; (3) to investigate if ISI has a larger effect on grade 3 student outcomes than for grade 1 or 2; and (4) to investigate if ISI can affect students who have previously received less effective literacy instruction (see Exhibits E10 through E13 for specific data). This study used multiple study designs, and the literature review team considered results from different designs separately.

First, there was a longitudinal design, which followed students from grades 1–3. The measures used for this portion of the study were the Letter-Word Identification and Passage Comprehension subtests from the Woodcock-Johnson III Tests of Achievement. The authors found that students who spent more time in intervention classrooms made larger gains on a standardized reading measure than comparison students. For this analysis, the authors created factor scores using Letter-Word Identification and Passage Comprehension subtest scores from the Woodcock-Johnson III. This portion of the study was reviewed as a QED because the students were not assigned *randomly* to the intervention and comparison conditions. Because the study did not clearly report baseline data, it was impossible to determine whether the groups were equivalent at baseline and therefore whether the intervention was responsible for the effects found. Author-reported findings should therefore be interpreted with caution.

Second, Connor et al. (2013) used a within-grade design for first-grade, second-grade, and third-grade effects. This design also used the Letter-Word Identification and Passage Comprehension subtests from the Woodcock-Johnson III. Authors followed a group of students from first grade to third grade and randomly assigned teachers to conditions at the start of each grade. The authors found that first-grade, second-grade, and third-grade students in the intervention condition scored significantly higher than their control condition counterparts in Letter-Word Identification and Passage Comprehension. Authors generated these results from well-designed cluster RCTs with low attrition, and the researchers accounted for the clustering of students within schools. Findings for the within-grade portion of the 2013 study can therefore be confidently attributed to the ISI and A2i bundled intervention. This design has the potential to meet WWC standards for strong casual evidence.

Small-Group Differentiated Instruction on Reading

One RCT study compared the strategy of grouping students by learning style preferences (i.e., visual, auditory, tactile, or kinesthetic) with the strategy of grouping students by pre-intervention reading achievement. There were no discernible effects between the two strategies.

Eastman (2010) conducted a student-level RCT with the goal of investigating reading instruction utilizing learning style preferences of first-grade students. Eastman defined learning style preferences as the ways in which learners prefer to approach learning tasks according to four categories: visual, auditory, tactile, or kinesthetic. She viewed the strategy of grouping students based on these preferences as a means to minimize the potential stigma of being in a "low-ability" group while still capitalizing on homogeneous small-group instruction. The intervention was conducted as an afterschool program for a group of 45 students in a Midwestern school. Intervention students were placed into groups based on these learning styles and provided with afterschool reading instruction, customized to their learning style. Comparison students were placed into groups based on reading level and received afterschool reading instruction not based on learning style. Achievement was measured using running record reading assessments to provide a total number of reading errors. With this assessment, students read a passage while teachers record miscues and errors in order to give insight into the students' reading strategies. The author reported that small-group reading instruction based on learning style had no discernable effect on reading achievement relative to small-group reading instruction based on reading level (see Exhibit E29). The literature review research team reviewed the RCT for evidence of effectiveness, during which a methodological issue was uncovered. The outcome was not a standardized test and therefore does not have established reliability or validity; the authors did not provide additional evidence related to reliability and validity. Therefore, findings should be interpreted with caution.

In three studies (two QEDS and one pre-test/post-test design), the instructional approaches placed students in homogeneous groups based on their reading achievement. Authors reported mixed effects of differentiated instruction for reading outcomes, analyzing students with higher and lower initial skills together.

In a study of differentiated instruction, Neel (2006) sought to assess the impact of small-group differentiated instruction on reading outcomes. In a QED, the author assigned students to small groups based on prior academic achievement and performance. In the intervention condition, grade 1 teachers provided one hour of supplemental small-group instruction to students on literacy comprehension, written composition, and word study and language. Grade 1 teachers in the comparison condition provided whole-class instruction on the same topics as usual, without students being placed into small groups based on achievement. The author measured outcomes using the Texas Primary Reading Inventory (TPRI) and the Developmental Reading Assessment (DRA). Neel (2006) found that the percentage of students who developed the ability to detect final sounds as measured by the TPRI at post-test was higher among intervention students (93 percent) than comparison students (82 percent). No other subscales of the TPRI yielded statistically significant results. There also was no statistically significant impact of the intervention on DRA student scores (see Exhibits E14 through E18). Neel concluded that, overall, contextually modified, developmentally appropriate literacy instruction in small groups did *not* produce statistically significant and higher achievement relative to the comparison condition.

Because Neel (2006) used a QED design, the research team reviewed the study for evidence of effectiveness, during which some methodological issues were uncovered. The author did not report the analytic sample's baseline data (sample sizes, means, and standard deviations) for testing whether the two groups were equivalent along these dimensions prior to the intervention. Instead, the authors provided baseline data for a different sample. Furthermore, all of the intervention students came from a single school and all comparison students came from a different school. This represents a confounding variable that makes it impossible to disentangle any treatment effects from the effects of belonging to the treatment school. For these reasons, the study findings should be interpreted with caution.

The goal of the study by Saylor (2008) was to evaluate the effect of differentiated instruction delivered through small ability groups on emergent literacy skills, including phonological, phonemic, and phonic skills. To do this, Saylor used a QED with students non-randomly assigned to the intervention and comparison conditions. In her study, kindergarten students in the intervention condition were divided into groups based on their areas of academic need, as determined by their scores on one general measure—the Georgia Kindergarten Assessment Program-Revised (GKAP-R)—and two literacy measures—the DIBELS and Basic Literacy Test. These instruments also served as the pre-test and post-test measures. For the intervention, teachers used differentiation strategies for literacy instruction for 60 minutes daily during language arts for three months. Comparison data came from the same three teachers' classrooms in the year before the intervention was implemented.

Saylor reported that the students in the intervention condition improved by an average of 13.49 points on the DIBELS Letter Naming Fluency (LNF) subscale, whereas the students in the comparison condition improved by an average of 6.0 points (see Exhibits E19 and E20). The difference was found to be statistically significant. There were no findings reported for the Initial Sound Fluency subscale of the DIBELS. The study had a methodological flaw in terms of standards for a well-designed QED; the use of a control group from a year prior to the treatment group year is considered methodologically inappropriate because time is a confounding factor that may have an effect on outcomes that cannot be eliminated by the study design. For this reason, the positive effect found for the LNF subscale of the DIBELS should be interpreted with caution.

In another study, Menzies, Mahdavi, and Lewis (2008) were interested in approaches to improve the reading performance of 42 grade 1 students from a small urban elementary school in southern California. The goal of the study was to assess whether student performance levels improved over time, and if improvement rates differed depending on students' initial skill level. Authors assessed achievement using the following measures: the DRA Test of Early Reading Ability–Revised (TERA-R), and DIBELS. Using a single-group pre-test/post-test design, they evaluated differentiated instructional practices by placing students into smaller groups according to their performance level (the authors labeled students as "at risk," "typically performing," and "proficient"). The first group (at risk) focused on phonemic awareness for students who struggled in this area. The second group (typically performing) emphasized decoding and fluency but did not include direct phonemic work. The third group (proficient) used guided-reading techniques that varied depending on the text. Teachers also received additional support through collaboration with other teachers and access to a literacy coach.

The authors found that students' post-test reading scores were higher and statistically significant compared with pre-test reading scores across all three groups. The authors also examined the scores for students in each of the groups separately; these findings are reported below in the subsequent section. For the main effect, the authors concluded that students achieved positive reading gains from experiencing small-group differentiated instruction and that 90 percent of the sample was reading at grade level by the end of the year. The authors also observed that gains were substantial relative to previous school years. The literature review research team did not review Menzies, Mahdavi, and Lewis (2008) for evidence of effectiveness because it did not include a comparison group. Data are not presented in Appendix E due to the design of the study. Findings should be interpreted with caution as the substantial gains in student reading scores may or may not be due to the intervention.

Three additional studies and one of the aforementioned studies analyzed ability grouping, examining results for students with different pre-intervention skills separately. Results from the three QED studies suggest that ability grouping can benefit students with higher initial reading skills, with less benefit to students with lower initial skills. One descriptive study suggests that students with medium- and lower-skill reading levels benefit, but students with higher initial skills do not.

A set of studies also tested the degree to which students in various achievement groups receive varying benefit from differentiated instruction. First, using data from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) data set and correlational analyses, Condron (2005, 2008) evaluated the effectiveness of skill-based grouping and curriculum differentiation. The author compared the reading improvement of first- and third-grade students whose teachers used skill-based grouping to differentiate instruction with students whose teachers did not use skill-based grouping. The ECLS-K sample used by Condron (2005) included 21,260 students who began kindergarten in the fall of 1998. Condron found that for this sample of students, low-ability first-grade students taught in homogeneous groups experienced less gain on reading outcomes (letter recognition, beginning sounds, ending sounds, sight comprehension of words, and comprehension of words in context) than a comparison group of students who were not taught in a classroom using ability grouping. Using the same sample, Condron (2008) found that, by third grade, students placed in the low-skill groups still gained fewer reading skills than their non-grouped peers; however, first- and third-grade students placed in high-skill groups demonstrated greater reading gains as compared with their non-grouped peers (see Exhibits E21 through E28). Because these studies used QED designs, they were reviewed for effectiveness. This review found that, in both studies, the author did not provide sufficient evidence that the groups were similar prior to the delivery of differentiated instruction. Without this evidence, it is difficult to determine whether the differences between groups on the outcome can be attributed to the intervention, pre-existing differences between the groups, or both. For this reason, the study findings should be interpreted with caution.

A study by Hong and colleagues (2012) also used the same sample of ECLS-K national data to investigate the impact of ability grouping on academic outcomes but did not include a non-grouped comparison condition. The authors' goal was to challenge the belief that homogeneous ability groupings benefit high-ability students at the expense of low-ability students. The researchers explored the effect of homogeneous ability grouping (at various student performance levels) on three outcomes: (1) students' literacy scores, (2) students' approaches to learning, and

(3) students' internalizing behavior problems. These outcomes were measured using a kindergarten literacy assessment, which the study authors did not describe in depth. The authors reported that students at all three ability levels (low, medium, and high) demonstrated similar growth on their overall literacy score. Within subdomains, students at different initial performance levels improved on different skills. For example, within the subdomains of sight words and comprehension of words in text, high-ability students demonstrated greater growth relative to medium- and low-ability students. Within the subdomain of learning beginning sounds and ending sounds, medium-ability students demonstrated greater growth than high- and low-ability students. Within the learning letter recognition subdomain, low-ability students demonstrated the greatest growth.

The authors also found that for the outcome of teacher reports of student approaches to learning, high-ability students performed the best, followed by medium-ability and then low-ability students. For the internalizing behavior problems outcome, the authors reported that low-ability students had the most internalizing problems, followed by medium- and then high-ability students. Data are not presented in Appendix E due to the design of the study (it did not include a comparison group). The study was not reviewed for evidence of effectiveness, and findings should be interpreted with caution.

As noted earlier, Menzies, Mahdavi, and Lewis (2008) also analyzed the effect of small-group differentiated instruction on students at different performance levels (low, typical, and proficient). The authors reported that students in the proficient group showed statistically significant growth from pre-test to post-test, as did the lowest performing group. However, the rate of growth for the lowest performing group of students was less than the typically performing group. The authors explained that this finding was likely due to their substantially lower pre-test scores. Again, because the authors' study did not include a comparison group, the findings cannot be confidently attributed to the intervention, and data are not presented in Appendix E.

Combining Individual and Small-Group Differentiated Instruction on Reading

One QED study found a greater percentage of growth in listening comprehension for students who received combined individual and small-group instruction relative to students who did not receive differentiated instruction.

A QED by Arnold (2008) investigated whether the Certified Learning Kindergarten (CLK) intervention had an impact on the academic development of kindergarten students, as measured by the Texas Primary Reading Inventory (TPRI). The CLK is an intervention that identifies a child's learning deficits and then customizes the curriculum to address those deficits. The CLK utilizes three different modes of instruction: group-oriented instruction, independent workbook instruction, and individual computer instruction. The program has a software management system that helps assign a student's daily schedule. By the end of the school year, student scores in the intervention condition grew 41 percent on the TPRI screening section compared with 40 percent growth for students in the comparison condition. In the listening comprehension section of the TPRI, CLK students showed 20 percent growth compared with 13 percent growth for comparison students (see Exhibit E30). The authors did not report statistical significance.

Because this was a QED, the research team reviewed the study for evidence of effectiveness, during which some methodological issues were uncovered. Specifically, the review found that groups were nonequivalent on listening comprehension before the intervention, and because the screening section falls into the same domain as listening comprehension, both analyses are considered nonequivalent at baseline. This means that a large portion of the difference found at post-test may have been due to the differences found prior to the intervention. For these reasons, findings from this study should be interpreted with caution.

Writing Programs and Practices

Three studies (one QED and two single-group pre-test/post-test designs) suggest that some students may benefit from collaborative, interactive writing sessions or from specific writing tools or prompts.

Roth (2009) conducted a QED to examine whether Interactive Writing, a dynamic and unscripted approach to writing instruction for primary grades, could improve the independent writing of first-grade students who attended low-income, urban public schools. In Interactive Writing, teachers work collaboratively with a student one-on-one to create a writing passage. Because teachers focus on individual students, they can customize their work with students based on individual needs. The intervention group was compared with a business-as-usual group. Roth measured writing improvement using two outcome measures: the Writing Samples subtest of the Woodcock-Johnson III and a researcher-developed writing prompt rubric containing 10 subscales. The writing prompt required students to respond to two prompts: (1) write and draw about something you do with your family and (2) write and draw a story about someone you know.

Roth found that Interactive Writing was an efficient daily practice to improve the quality of students' independent writing, although findings differed across the two measures used. Students in the Interactive Writing group outperformed students in the comparison group. Lower initial reading scores predicted greater gains in writing when assessed with the writing prompt; however, higher initial reading scores predicted greater gains in writing when assessed with the Writing Samples subtest (see Exhibit E31). Because the authors used a QED design, the research team reviewed the study for evidence of effectiveness, during which some methodological issues were uncovered. Although Roth tested for baseline equivalence between the groups, equivalence was not appropriately established for the outcome measure, making it impossible to conclude whether the impact was due to the intervention or to some other underlying difference between the two conditions. Therefore, all findings should be interpreted with caution.

Geisler and colleagues (2009) examined the effects of differentiating instruction for a group of five high-performing African-American students in a split first-/second-grade classroom. The authors described that the five students in the classroom are the only ones receiving the instructional strategies in order to build on their higher skills; in this way, those strategies represent differentiated instruction relative to the regular instruction that the remaining students in the classroom are receiving. However, the study did not collect data for a comparison group (e.g., a "business-as-usual" condition in which students did not receive the differentiated instruction). The two instructional strategies used with the high-achieving group were "self-counting" and a "synonym list," and were presented in a two-part intervention. Self-counting involved students counting the words they wrote in their writing samples after each writing

session, assisted by the teacher or researcher, and recording the number of total words as well as the number of different words written during the first three minutes of the writing session. This strategy was designed to help students develop an awareness of their writing output in order to encourage them to write more. In the second part of the intervention, teachers added use of the synonym list, which involved giving the students a list of synonyms for words most commonly used in first-grade writing (e.g., "big" can be replaced by "large," "huge," "enormous," "giant," and "gigantic") to encourage students to use more complex words in their writing.

The authors found that, on average, the number of different words used increased from pre-test to post-test during the first part of the intervention (self-counting only). All five students also increased in their number of total words written during the first phase. During the second phase, the introduction of the synonym list, one student decreased in the number of total words and one student increased only slightly. The other three students increased more substantially in their total number of words. The authors used a final outcome measure, Generalization Probes, which involved students completing a writing task without explicit use of the two strategies. The authors found that all five students wrote a greater number of total words and a greater number of different words in each successive Generalization Probe. The authors suggest that the skills students learned during the intervention sessions were beginning to generalize to overall writing performance. There are multiple methodological issues with this study. Because of the extremely small sample size used in this study and the lack of a comparison group, it is impossible to confidently attribute the findings to the intervention; therefore, all findings should be interpreted with caution. Data are not presented in Appendix E.

In another single-group pre-test/post-test design study, Case-Smith and colleagues (2011) examined the impact of Write Start, a handwriting intervention, on handwriting legibility as well as speed and writing fluency. This intervention used a co-teaching model in which occupational therapists and teachers collaborated to develop and implement Write Start, a 12-week classroom-embedded intervention for first graders, with particular attention paid to individual students' needs. The teaching staff, therefore, conducted differentiation at the individual level during each writing session. Handwriting legibility and speed were assessed using the Evaluation Tool of Children's Handwriting–Manuscript and the Minnesota Handwriting Assessment. Writing performance was measured using the Writing Fluency and Writing Samples subtests from the Woodcock-Johnson III.

During the 12-week period of the intervention, students' legibility scores progressed from a mean of 62 percent to 87 percent. The score of 87 percent indicated that, on average, students achieved legible handwriting that an audience can read without effort. On the six-month follow-up, measurement legibility was maintained. The students also made improvements in handwriting speed; the average time required to write the alphabet decreased from greater than 3 minutes to 1 1/2 minutes. Case-Smith et al. (2011) concluded that when Write Start is implemented with high fidelity by trained occupational therapists and teachers, it can lead to significant gains in handwriting legibility, speed, and writing fluency. Because this study used a single-group pre-test/post-test design, the research team did not review it for evidence of effectiveness. Also, because the study did not have a comparison group, it is not possible to rule out alternative explanations for the observed gains on writing measures. Data are not presented in Appendix E.

Differentiated Instruction Strategies in Qualitative Case Studies

In addition, four qualitative studies met criteria for the literature review. These studies focused almost entirely on processes and strategies for implementing differentiated instruction for mathematics instruction and on researcher perceptions (who, in two practitioner inquiry studies, were the teachers themselves) of factors that facilitate or hinder implementation. None of these studies reported student outcome data; therefore, the studies are used here to provide additional insight into differentiated instruction implementation rather than as evidence of effects.

Opportunities for peer collaboration and guidance by mentors, such as coaches, may be helpful to improve teacher practice related to differentiation.

The four math studies focused on practices at the kindergarten (Bofferding, Kemmerle, & Murata, 2012; Ensign, 2012), first-grade (Holden, 2007), or combined first- and second-grade (Kobelin, 2009) level. In two of the studies, researchers and coaches facilitated differentiated instruction by bringing teachers together to collaborate and share practices. In the first study, Bofferding, Kemmerle, and Murata (2012) focused on three kindergarten teachers who engaged in a lesson study approach in which the teachers (a) met together to plan a lesson, (b) observed each other's teaching, and (c) reflected together on student learning. The study reported on approaches to teaching students particular math standards relating to students' understanding of part-whole relations in combining numbers to make 10. The teachers met four times to consider their kindergarten students' current level of thinking and to plan aspects of instruction that might need to be individualized or differentiated among students. Each teacher eventually taught the content, with the others observing, using instructional materials and strategies to allow students to access problems in different ways. Some students used concrete manipulatives to solve problems; others used an activity sheet that guided students to keep them on task by limiting their exploration of number concepts to only 10. Students who quickly found one solution to problems were challenged to find all solutions and to be more strategic in their approach. Study authors reported that the lesson study approach, which purports to develop teachers' "researcher lens" on their own practice (Choksi & Fernandez, 2004, cited in Bofferding, Kemmerle, & Murata, 2012), was successful in helping teachers better understand and tailor instruction for individual student thinking.

Ensign (2012) reported on a single kindergarten teacher's practices related to differentiation in the context of a district that funded school- and district-based math coaches as well as increased math professional development funding for coaches and teachers. Developing effective differentiated teaching strategies, and allowing teachers to observe colleagues who competently instruct at multiple levels, became a key component of the coaches' work. The kindergarten teacher featured in the study developed a choice system to ensure that all students were actively engaged. Following a short whole-group lesson, students chose from an array of math games and activities that focused on various math standards. While students worked on their own and at their own level, the teacher focused on instructing individuals or small groups and conducted performance assessments. As part of the broader coaching and professional development initiative in the district, video of model teachers was used for professional development by coaches. The broader initiative allowed teachers release time to observe and debrief with model teachers and to attend intensive trainings on differentiation, as well as professional development hours to work with a coach to develop differentiation strategies in their own classrooms.

Teachers formed a book club to read and discuss strategies for differentiation. The authors conclude that coaches played a critical role in increasing and improving differentiated instruction for math.

In the other two math studies, the authors reported on trial-and-error attempts at implementing differentiated instruction in their own classrooms. In one study (Holden, 2007), the teacher-researcher implemented differentiated instruction in her own first-grade classroom by providing flexible math problems in which blanks can be filled in to adapt problems to varying levels of difficulty and using different math strategies. The teacher created a form to document the problems students solved and their strategies. The teacher reflected that the approach can be used to better scaffold students, or to move students progressively toward stronger understanding, in their problem solving. According to the author, the approach helped record growth over time and organize small groups of students working at the same level.

In another study (Kobelin, 2009), the teacher-researcher learned to implement differentiated approaches because there were varying skills in her mixed-age first- and second-grade classroom. As one method for differentiation, the teacher used open-ended tasks that have no single answer or method to determine an answer. Working independently or with the teacher or peers, students can be challenged to solve problems in more than one way and to find more complex solutions. As a second method, the teacher used student-paced, tiered tasks developed to address multiple, specific skill levels. In the last method, the teacher planned "spiraling-scaffolded tasks" in which students at different levels or different grades address the same concept (e.g., time) but at varying levels of complexity and with different teacher modeling or coaching. Like the teacher in Ensign (2012), the teacher in Kobelin (2009) utilized a combination of short, whole-group lessons with subsequent instruction periods involving student choice and independence with math activities, especially for those who were more advanced. The teacher-researcher emphasized that math is very challenging to differentiate, that differentiation in math is less common compared with reading, and that publishers of math curricular materials do not generally provide plans and materials to facilitate differentiation; therefore, she needed to learn effective differentiation practices through experimentation.

Conclusion

Overall, the findings from the 17 quantitative studies in the literature review suggest that differentiated instruction delivered individually or in small ability-based groups may have an impact on reading and writing outcomes for students in kindergarten and first grade. It is critical to note that, based on the information in the published studies, only one of the 17 studies from the quantitative study pool had the potential to meet evidence standards as a well-designed and well-implemented RCT (the within-grade, first-grade study published in Connor et al., 2013). This suggests that further research on differentiated instruction interventions for early elementary students would be strengthened by more rigorous RCTs and QEDs that are careful to establish baseline equivalence between the intervention and comparison groups and, for RCTs, are vigilant about reporting attrition data.

The four qualitative studies provide information about processes and strategies for stakeholders who may seek to implement differentiated instruction for mathematics, but these studies do not provide evidence of effects. The qualitative studies suggest that differentiated instruction may be

difficult to implement and requires careful planning and reflection on the part of teachers. Opportunity and time for teachers to carefully plan, reflect, and collaborate with peers on differentiated instruction practice and to receive guidance by mentors, such as coaches, may be helpful to improve teacher practice related to differentiation. These implementation recommendations have not been empirically validated and therefore require further research.

IV. Conclusion

Reflecting the state of the research in the field, key findings for preschool and K–3 alignment focus on theoretical and policy considerations, while findings for differentiated instruction summarize the results of quantitative studies.

Summary of P–3 Alignment Findings

The overarching goal of P–3 alignment policy and practices is to provide a coherent educational experience for students as they progress from preschool through elementary school that could potentially sustain the positive effects of preschool. Findings from the literature—including (1) 49 policy or theory resources; (2) nine qualitative studies, most of which used a case study approach; (3) two quantitative studies; and (4) one mixed-methods study—reveal a widely held recommendation for alignment of standards, curriculum, instruction, assessments, and environments across preschool and grades K–3 as an approach for providing high-quality education to students in this grade range. Currently, there is very little extant research that empirically supports this recommendation, suggesting that outcomes of P–3 alignment initiatives require further research.

The literature provides recommendations for stakeholders who seek to implement and design P−3 initiatives. These include the following:

- Consider establishing similar teacher education and training requirements across preschool and elementary education job positions.
- Create systems that educators can use to better tailor instruction to meet students' needs. Such systems would link individual student data from public and private early childhood programs to students' public school data. Support implementation of P−3 initiatives through the management practices district administrators put in place. Administrators should consider:
    - Involving early childhood education providers and grade K–3 teachers in planning P–3 initiatives.
    - Establishing procedures to ensure implementation fidelity of P–3 elements.
    - Specifying measurable student achievement benchmarks.
    - Holding principals and teachers accountable for achieving benchmarks.
- Consider and find solutions to challenges to P−3 initiatives, including: policies that inhibit the blending of various sources of funding, instability of preschool funding, resistance by practitioners to integration of preschool and the K–3 grades, and the organization of elementary education classrooms, buildings, and enrollment.

These implementation recommendations also have not been empirically validated and therefore require further research.

Summary of Differentiated Instruction Findings

Educators propose that differentiated instruction is a way to meet students' diverse needs by having teachers deliver instruction through different means or customizing instruction for different performance levels so that all students have access to instruction that will address and match their various skills, motivation, interests, and/or backgrounds. In terms of sustaining the effects of preschool, authors suggest that differentiated instruction may be a way to maintain the growth rate or learning trajectory of children who make early gains in preschool as they enter elementary school (e.g., Kauerz, 2006; Tomlinson et al., 2003).

Overall, the findings from the 17 quantitative studies in the literature review suggest that differentiated instruction delivered individually or in small ability-based groups may have an impact on reading and writing outcomes for students in kindergarten and first grade. Furthermore, the evidence suggests that potential effects may differ depending on the pre-intervention skills of the students, in particular for differentiated instruction delivered in small groups. The evidence shows mixed results—two studies suggest that ability grouping can benefit students with higher initial reading skills, with less benefit to students with lower initial skills. One study suggests that students with medium- and lower-skill reading levels benefit, but students with higher initial skills do not. It is critical to note that only one of the 17 studies from the quantitative study pool had the potential to meet WWC evidence standards as a well-designed and well-implemented RCT (the within-grade, first-grade study published in Connor et al., 2013). This suggests that differentiated instruction interventions for early elementary students require further research that uses well-designed RCTs and QEDs. For example, RCTs should be careful to report sample sizes for attrition calculations and both RCTs with high attrition and QEDs should be careful to establish baseline equivalence between the intervention and comparison groups' analysis sample.

The four qualitative studies provide information about processes and strategies for stakeholders who may seek to implement differentiated instruction for mathematics, but these studies do not provide evidence of effects. The set of very small qualitative studies suggest that differentiated instruction may be difficult to implement and requires careful planning and reflection on the part of teachers. Opportunity and time for teachers to carefully plan, reflect, and collaborate with peers on differentiated instruction practice and to receive guidance by mentors, such as coaches, may be helpful to improve teacher practice related to differentiation. These implementation recommendations have not been empirically validated and therefore require further research.

References

ABCs of early education: Listening, asking, sharing, engaging. (2013). *American Teacher, 97*(5), 6–7.

10 action steps. (2011). *Principal, 91*(2), 9.

Advocates for Children of New Jersey. (2010). *Linking learning: The PreK–3rd path to school success.* Newark, NJ: Author.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *Elementary School Journal, 111*(4), 535–560.

Anderson, K. M., (2007). Differentiating instruction to include all students. *Preventing School Failure, 51*(3), 49–54.

Andrews, R. J., Jargowsky, P., & Kuhne, K. (2012). *The effects of Texas's pre-kindergarten program on academic performance* (CALDER Working Paper No. 84). Washington, DC: National Center for Analysis of Longitudinal Data on Educational Research.

Archambault, F. X., Westberg, K. L., Brown, S. W., Hallmark, B. W., Zhang, W., & Emmons, C. L. (1993). Classroom practices used with gifted third and fourth grade students. *Journal for the Education of the Gifted, 16*(2), 103–119.

Arnold, A. M. P. (2008). The influence of respecting the individual child's learning system on early academic development. *ProQuest Dissertations and Theses*, 89250452.

Barnett, W. S. (2008). *Preschool education and its lasting effects: Research and policy implications.* Boulder, CO, & Tempe, AZ: Education and the Public Interest Center & Education Policy Research Unit. Retrieved from http://nepc.colorado.edu/files/PB-Barnett-EARLY-ED_FINAL.pdf

Bofferding, L., Kemmerle, M., & Murata, A. (2012). Making 10 my way. *Teaching Children Mathematics, 19*(3), 164–173.

Bogard, K. L. (2006). Testing a P–3 approach to education in three urban public schools. *Dissertation Abstracts International, 62,* A1589044.

Bogard, K., & Takanishi, R. (2005). P–3: An aligned and coordinated approach to education for children 3 to 8 years old. *Social Policy Report, 19*(3), 1–23.

Brett, J., Staniszewska, S., Newburn, M., Jones, N., & Taylor, L. (2011). A systematic mapping review of effective interventions for communicating with, supporting and providing information to parents of preterm infants. *BMJ Open, 1,* 1–11. Retrieved from http://bmjopen.bmj.com/content/1/1/e000023.full?sid=70e87115-3b80-4c26-8a42-fe26c7a60928

Brimijoin, K. (2005). Differentiation and high-stakes testing: An oxymoron? *Theory Into Practice*, *44*(3), 254–261.

Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report, 17*(1), 3–14.

Brown, B. V., & Bogard, K. (2007). Pre-kindergarten to 3rd grade (P–3) school-based resources and third grade outcomes. *CrossCurrents, 5* (2007-27). Washington, DC: Child Trends.

Buenafe, A. Q. (2011). *Pre-K as a school turnaround strategy.* Washington, DC: Pew Center.

Case-Smith, J., Holland, T., & Bishop, B. (2011). Effectiveness of an integrated handwriting program for first-grade students: A pilot study. *American Journal of Occupational Therapy, 65*(6), 670–678.

Center for Applied Research and Educational Improvement. (2013). *Education & learning program: PreK–3 professional development*. Minneapolis, MN: University of Minnesota, College of Education & Human Development.

Center for the Study of Educational Policy. (2012). *From birth to graduation and beyond: Aligning best practices in educational systems to improve learning outcomes.* Normal, IL: Illinois State University.

Claessens, A., Engel, M., & Curran, F. C. (2013). Academic content, student learning, and the persistence of preschool effects. *American Educational Research Journal*, 0002831213513634.

Clarke-Stewart, K. A., Vandell, D. L., Burchinal, M., O'Brien, M., & McCartney, K. (2002). Do regulable features of child-care homes affect children's development? *Early Childhood Research Quarterly, 17*(1), 52–86.

Clements, M. A., Reynolds, A. J., & Hickey, E. (2004). Site-level predictors of children's school and social competence in the Chicago Child-Parent Centers. *Early Childhood Research Quarterly, 19*, 273–296.

Cochran-Smith, M., & Lytle, S. L. (1999a). The teacher research movement: A decade later. *Educational Researcher, 28*(7), 15–25.

Cochran-Smith, M., & Lytle, S. L. (1999b). Relationships of knowledge and practice: Teacher learning in communities. *Review of Research in Education, 24,* 249–305.

Committee for Economic Development. (2012). *Unfinished business: Continued investment in child care and early education is critical to business and America's future.* Washington, DC: Author.

Condron, D. J. (2005). Stratification, skill grouping, and learning to read in first grade. *ProQuest Dissertations and Theses,* 305399562.

Condron, D. J. (2008). An early start: Skill grouping and unequal reading gains in the elementary years. *The Sociological Quarterly, 49,* 363–394.

Connor, C. M., Morrison, F. J., Fishman, B., Crowe, E. C., Al Otaiba, S. A., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science, 24*(8), 1408–1419.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science, 315,* 464–465.

Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J. R., Lundblom, E., Crowe, E. C., & Fishman, B. (2011). Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness, 4*(3), 173–207.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E.,…Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Development, 80*(1), 77–100.

Connor, C. M., Ponitz, C. C., Phillips, B. M., Travis, M. Q., Glasney, S., & Morrison, F. J. (2010). First graders' literacy and self-regulation gains: The effect of individualizing student instruction. *Journal of School Psychology, 48*(5), 433–455.

Cooper, H. M. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: Sage.

Demanchick, S. P., Peabody, M. A., & Johnson, D. B. (2009). Primary project: Fifty years of facilitating school adjustment. *Blending play therapy with cognitive behavioral therapy: Evidence-based and other effective treatments and techniques* (pp. 219–235). Hoboken, NJ: Wiley.

Donovan, L. F. (2010). *Linking learning: Building quality kindergarten in New Jersey.* Newark, NJ: Advocates for Children of New Jersey.

Eastman, V. L. (2010). Small group instruction: Reading instruction utilizing learning style preferences and the reading achievement of first grade students. *ProQuest Dissertations and Theses,* 305186345.

Ensign, J. (2012). Teacher-initiated differentiation. *Teaching Children Mathematics, 19*(3), 158–163.

EPPI Centre. (2010). *EPPI-Centre methods for conducting systematic reviews.* Retrieved from http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=hQBu8y4uVwI%3D&tabid=88

Frede, E., Jung, K., Barnett, W. S., Lamy, C. E., & Figueras, A. (2007). *The Abbott Preschool Program Longitudinal Effects Study (APPLES).* New Brunswick, NJ: National Institute for Early Education Research.

Garland, S. (2011). Pre-K–grade 3 continuum gets sharper focus. *Education Week, 30*(22), 7–7.

Gates Foundation. (2011). *Early learning in Washington state.* Seattle, WA: Author.

Geisler, J. L., Hessler, T., Gardner, R., & Lovelace, T. S. (2009). Differentiated writing interventions for high-achieving urban African American elementary students. *Journal of Advanced Academics, 20*(2), 214–247.

Gettinger, M., & Stoiber, K. C. (2012). Curriculum-based early literacy assessment and differentiated instruction with high-risk preschoolers. *Reading Psychology, 33*(1–2), 11–46.

Gilliam, W. S., & Zigler, E. F. (2001). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly, 15*(4), 441–473.

Goldstein, L. S., & Bauml, M. (2012). Strategies and suggestions for pre-K-grade 3 teachers in public school contexts. *Young Children, 67*(3), 96–103.

Gormley, W. T., & Phillips, D. (2005). *The effects of universal pre-K in Oklahoma: Research highlights and policy implications.* Washington, DC: Georgetown University. Retrieved from http://www.crocus.georgetown.edu/reports/CROCUSworkingpaper2.pdf

Grantmakers for Education. (2006). *Making the most of our investments: How P–3 alignment can close the achievement gap from the start.* Portland, OR: Grantmakers for Education.

Groark, C. J., Mehaffie, K. E., McCall, R. B., & Greenberg, M. T. (Eds.) (2007). *Evidence-based practices and programs for early childhood care and education.* Thousand Oaks, CA: Corwin.

Guernsey, L., Bornfreund, L., McCann, C., & Williams, C. (2014). *Subprime learning: Early education in America since the great recession.* Washington, DC: New American Foundation.

Halpern, R. (2013). Tying early childhood education more closely to schooling: Promise, perils and practical problems. *Teachers College Record, 115*(1), 1–28.

Hays, P. A. (2004). Case study research. In K. deMarrais & S. D. Lapan (Eds.), *Foundations for research: Methods of inquiry in education and the social sciences* (pp. 217–234). Mahwah, NJ: Lawrence Erlbaum Associates.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the High Scope Perry Preschool Program. *Journal of Public Economics*, *94*(1), 114–128.

Hernandez, D. J. (2012). *PreK–3rd: Next steps for state longitudinal data systems. PreK-3rd policy action brief no. eight.* New York, NY: Foundation for Child Development.

Holden, B. (2007). Preparing for problem solving. *Teaching Children Mathematics, 14*(5), 290–295.

Hong, G., Corter, C., Hong, Y., & Pelletier, J. (2012). Differential effects of literacy instruction time and homogeneous ability grouping in kindergarten classrooms: Who will benefit? Who will suffer? *Educational Evaluation and Policy Analysis, 34*, 69–88.

Howard, M. (2008). *Early care and education: The progress of education reform* (Vol. 9, No. 1). Denver, CO: Education Commission of the States.

Howes, C. (1988). Relations between early child care and schooling. *Developmental Psychology, 24*(1), 53.

Human Capital Research Collaborative. (2014a). *2013–2014 Midwest Child-Parent Center expansion: Program guidelines and requirements.* Minneapolis, MN: University of Minnesota.

Human Capital Research Collaborative. (2014b). *Child-Parent Centers fact briefs: Demystifying "fade-out" effects.* Minneapolis, MN: Author.

Jacobson, L. (2009). *On the cusp in California: How PreK–3rd strategies could improve education in the golden state.* Washington, DC: New America Foundation.

Jacobson, R., Jacobson, L., & Blank, M. J. (2012). *Building blocks: An examination of the collaborative approach community schools are using to bolster early childhood development.* Washington, DC: Coalition for Community Schools, Institute for Educational Leadership.

Jones, R. E., Yssel, N., & Grant, C. (2012). Reading instruction in tier 1: Bridging the gaps by nesting evidence-based interventions within differentiated instruction. *Psychology in the Schools, 49*(3), 210–218.

Kagan, S. L., Carroll, J., Comer, J. P., & Scott-Little, C. (2006). Alignment: A missing link in early childhood transitions? *Young Children, 61*(5), 26–32.

Kagan, S. L., & Kauerz, K. (2010). Governance and transition. In S.L. Kagan & K. Tarrant (Eds.), *Transitions for young children: Creating connections across early childhood systems* (pp. 243–265). Baltimore, MD: Brookes.

Karoly, L. A., & Bigelow, J. H. (2005). *The economics of investing in universal preschool education in California*. Santa Monica, CA: RAND.

Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2005). *Early childhood interventions: Proven results, future promise*. Santa Monica, CA: RAND. Retrieved from http://www.rand.org/pubs/monographs/MG341

Kauerz, K. (2006). *Ladders of learning: Fighting fade-out by advancing P–3 alignment.* Washington, DC: New America Foundation Early Education Initiative.

Kauerz, K. A. (2009). The early childhood and elementary education continuum: Constructing an understanding of P–3 as state-level policy reform. *ProQuest Dissertations and Theses*, 304870194.

Kauerz, K., & Coffman, J. (2013). *Framework for planning, implementing, and evaluating PreK–3rd grade approaches.* Seattle, WA: University of Washington, College of Education.

King, J. (2006). *Closing the achievement gap through expanded access to quality early education in grades P–3.* Washington, DC: New America Foundation Early Education Initiative.

Kobelin, M. (2009). Multi-age made me do it: A teacher tackles differentiation in math instruction. *Schools: Studies in Education, 6*(1), 10–22.

Lawrence-Brown, D. (2004). Differentiated instruction: Inclusive strategies for standards-based learning that benefit the whole class. *American Secondary Education*, 34–62.

Lee, V. E., & Loeb, S. (1995). Where do Head Start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis, 17*(1), 62–82.

Lesaux, N. (2010). *Turning the page: Refocusing Massachusetts for reading success.* Springfield, MA: Strategies for Children.

Lipsey, M. W., Farran, D. C., Hofer, K. G. (2015). *A randomized control trial of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade.* Nashville, TN: Vanderbilt University, Peabody Research Institute.

Maeroff, G. I. (2003). *Universal pre-kindergarten: State of play* (Working Paper Series). New York, NY: Foundation for Child Development.

Magnuson, K., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2005). Inequality in children's school readiness and public funding. *Focus, 24*(1), 12–18.

Manship, K., Madsen, S., Mezzanotte, J., & Fain, G. (2013). *Evaluation of the stretch to kindergarten program: 2012 findings.* San Mateo, CA: American Institutes for Research.

Marietta, G. (2010a). *Lessons for PreK–3rd from Montgomery County Public Schools: An FCD case study.* New York, NY: Foundation for Child Development.

Marietta, G. (2010b). *PreK–3rd: How superintendents lead change. PreK–3rd policy action brief, No. 5.* New York, NY: Foundation for Child Development.

Marietta, G., & Marietta, S. H. (2013a). *PreK–3rd's lasting architecture: Successfully serving linguistically and culturally diverse students in Union City, New Jersey.* New York, NY: Foundation for Child Development.

Marietta, G., & Marietta, S. H. (2013b). *The promise of PreK–3rd: Promoting academic excellence for dual language learners in Red Bank Public Schools.* New York, NY: Foundation for Child Development.

Mead, S. (2009). *Education reform starts early: Lessons from New Jersey: PreK–3rd reform efforts.* Washington, DC: New America Foundation.

Menzies, H. M., Mahdavi, J. N., & Lewis, J. L. (2008). Early intervention in reading: From research to practice. *Remedial & Special Education, 29*(2), 67–77.

Metcalf, T. (2013). What's your plan? Accurate decision making within a multi-tier system of supports: Critical areas in tier 1. *RTI Action Network.*

Murata, A. (2013). Diversity and high academic expectations without tracking: Inclusively responsive instruction. *Journal of the Learning Sciences, 22*(2), 312–335.

NALEO Education Leadership Initiative. (2008). *Latino legislative hearing on pre–K & the early grades* Washington, DC: Author.

National Association of Elementary School Principals. (2011). *Building & supporting an aligned system: A vision for transforming education across the pre-K–grade three years.* Alexandria, VA: Author.

National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network [ECCRN]. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science, 4*(3), 116–135.

Neel, J. L. (2006). The effects of differentiated developmentally appropriate instruction of first grade learners. *ProQuest Dissertations and Theses,* 304912978.

New, R., Palsha, S., & Ritchie, S. (2009). *Issues in PreK–3rd education: A FirstSchool framework for curriculum and instruction* (Issue 7). Chapel Hill, NC: Frank Porter Graham Child Development Institute.

Nyhan, P. (2011). *The power of PreK–3rd: How a small foundation helped push Washington State to the forefront of the PreK–3rd movement.* New York, NY: Foundation for Child Development.

Parsons, S. A. (2012). Adaptive teaching in literacy instruction: Case studies of two teachers. *Journal of Literacy Research, 44,* 149–170.

Parsons, S. A., Dodman, S. L., & Burrowbridge, S. C. (2013). Broadening the view of differentiated instruction: differentiation shouldn't end with planning but should continue as teachers adapt their instruction during lessons. *Phi Delta Kappan, 95*(1), 38.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72*(5), 1534–1553.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practice guide.* Oxford, UK: Blackwell Publishing.

Pritchard, I. A. (2002). Travelers and trolls: Practitioner research and institutional review boards. *Educational Researcher, 31*(3), 3–13.

QSR International. (2012). NVivo 10 [computer software]. Doncaster, Australia: Author.

Ramey, S. L., Ramey C. T., Phillips, M. M., Lanzi, R. G., Brezausek, C., Katholi, C. R., & Snyder, S. (2000). *Head Start children's entry into public school: A report on the National Head Start/Public School Early Childhood Transition Demonstration Study.* Birmingham, AL: University of Alabama at Birmingham, Civitan International Research Center.

Ravitch, S. M. (2014). The transformative power of taking an inquiry stance on practice: Practitioner research as narrative and counter-narrative. *Penn GSE Perspectives on Urban Education, 11*(1), 5–10.

Reynolds, A. J. (1993). One year of preschool intervention or two: Does it matter? *Early Childhood Research Quarterly, 10,* 1–33.

Reynolds, A., Magnuson, K., & Ou, S. (2006). *P–3 education: Programs and practices that work in children's first decade.* New York, NY: Foundation for Child Development.

Reynolds, A. J., Magnuson, K. A., & Ou, S. (2010). Preschool-to-third grade programs and practices: A review of research. *Children and Youth Services Review, 32,* 1121–1131.

Reynolds, A. J., & Ou, S. (2011). Paths of effects from preschool to adult well-being: A confirmatory analysis of the Child-Parent Center Program. *Child Development, 82*(2), 555–582.

Reynolds, A. J., & Temple, J. A. (2008). Cost-effective early childhood development programs from preschool to third grade. *Annual Review of Clinical Psychology, 4,* 109–139.

Reynolds, A. J., Temple, J. A., Ou, S., Robertson, D. L., Mersky, J. P., Topitzes, J. W., & Niles, M. D. (2007). Effects of a school-based early childhood intervention on adult health and well-being: A 19-year follow-up of low-income families. *Archives of Pediatric and Adolescent Medicine, 161*(8), 730–739.

Rice, C. (2007). *Building strong rungs to build sturdy ladders: The status of preschool–3rd grade systems in New Jersey.* Newark, NJ: Association for Children of New Jersey.

Rice, C. (2008a). *Developing an advocacy strategy for New Jersey's P–3 agenda inside and out.* Newark, NJ: Association for Children of New Jersey.

Rice, C. (2008b). *Seizing the opportunity: Building P–3 systems in New Jersey's school districts.* Newark, NJ: Association for Children of New Jersey.

Rice, C. (2010). *Linking learning: Congress should follow New Jersey's lead on early learning.* Newark, NJ: Advocates for Children of New Jersey.

Ritchie, S., Clifford, R. M., Malloy, W. W., Cobb, C. T., & Crawford, G. M. (2010). Ready or not? Schools' readiness for young children. In S. L. Kagan & K. Tarrant (Eds.), *Transitions for young children: Creating connections across early childhood systems* (pp. 161–184). Baltimore, MD: Brookes.

Ritchie, S., Maxwell, K., & Clifford, R. (2009). What is FirstSchool? *Issues in PreK–3rd Education* (Issue 1). Chapel Hill, NC: Frank Porter Graham Child Development Institute.

Ritchie, S., Maxwell, K., & Clifford, R. M. (2007). FirstSchool: A new vision for education. In R. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 85–96). Baltimore, MD: Brookes.

Roth, K. (2009). Interactive writing: Investigating the effectiveness of a dynamic approach to writing instruction for first graders. *ProQuest Dissertations and Theses,* 304894167.

Saylor, A. R. (2008). The impact of differentiated instruction on emergent literacy skills with kindergarten students. *ProQuest Dissertations and Theses,* 304390344.

Scott-Little, C., & Reid, J. (2010). Aligning the content of early childhood care and education to promote effective transitions. In S. L. Kagan & K. Tarrant (Eds.), *Transitions in the early years: Creating a system on continuity* (pp. 109–133). Baltimore, MD: Brookes.

Severns, M. (2012). *Starting early with English language learners: First lessons from Illinois.* Washington, DC: New America Foundation.

Shore, R. (2009). *The case for investing in preK–3rd education: Challenging myths about school reform.* New York, NY: Foundation for Child Development.

Stanford, B., & Reeves, S. (2009). Making it happen: Using differentiated instruction, retrofit framework, and universal design for learning. *Teaching Exceptional Children Plus*, *5*(6), 6.

Takanishi, R. (2010). Children of 2020: Creating a better tomorrow. In V. Washington & J. Andrews (Eds.), *Pre-K–third grade: A paradigm shift* (pp. 28–31). Washington, DC: Council for Professional Recognition.

Takanishi, R., & Bogard, K. L. (2007). Effective educational programs for young children: What we need to know. *Child Development Perspectives, 1*(1), 40–45.

Takanishi, R., & Kauerz, K. (2008). P inclusion: Getting serious about a P–16 education system. *Phi Delta Kappan, 89,* 480–487.

The Pre-K Coalition. (2011a). *Ensuring America's future: Policy statements and recommendations from national education organizations.* Washington, DC: Author.

The Pre-K Coalition. (2011b). *The importance of aligning pre-K through 3rd grade.* Washington, DC: Author.

Tomlinson, C. A. (1999). *The differentiated classroom: Responding to the needs of all learners.* Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. A. (2000). Differentiation of instruction in the elementary grades. *ERIC Digest*. Syracuse, NY: ERIC Clearinghouse on Elementary and Early Childhood Education.

Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). Alexandria, VA: ASCD.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K.,...Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, *27*(2–3), 119–145.

Tout, K., Halle, T., Daily, S., Albertson-Junkans, L., & Moodie, S. (2013). *The research base for a birth through age eight state policy framework* (2013-42). Bethesda, MD: Child Trends.

University of Chicago, Urban Education Institute, & Ounce of Prevention Fund. (2012). *Building a birth-to-college model: Professional learning communities.* New York, NY: Authors.

U.S. Department of Health and Human Services, Administration for Children and Families. (2010, January). *Head Start Impact Study: Final report*. Washington, DC: Author.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112–2130.

Westberg, K. L., Archambault, F. X., Dobyns, S. M., & Salvin, T. J. (1993). The classroom practices observation study. *Journal for the Education of the Gifted*, *16*(2), 120–146.

What Works Clearinghouse. (2010a). *Procedures and standards handbook, version 3.0.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf

What Works Clearinghouse. (2010b). *WWC single study review protocol for group design studies that estimate average intervention effects, version 2.0*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/ssr_protocol_v2.pdf

Zellman, G. L., & Kilburn, R. M. (2011). *Hawai'i P–3 initiative: Findings from the first year of the evaluation.* Santa Monica, CA: RAND.

APPENDIX A. SPECIFICATIONS FOR THE LITERATURE SEARCH

Appendix A.
Specifications for the Literature Search

Appendix A provides additional details on the parameters of the literature search to complement information presented in the text of the report.

Electronic Databases

The research team used the following core list of electronic databases to search for both topics:

1. *Academic Search Premier*
2. *Dissertation Abstracts*
3. *EconLit*
4. *Education Full Text*
5. *Education Resources Information Center (ERIC)*
6. *JSTOR*
7. *Professional Development Collection*
8. *PsycINFO*
9. *Sociological Abstracts*

Search Terms

The research team utilized the following terms for searches on the P–3 alignment topic:

"pre-K-grade three" OR "pre-K through third" OR "PreK-3rd" OR "P–3" OR "Pre-K-3rd" OR "ages 3 through 8" OR "ages 3-8" OR "age 3 to age 8" OR "pre-kindergarten through third grade" OR "pre-kindergarten through grade three" OR "preschool through third grade" OR "preschool through grade three" OR "preschool-grade three" OR "preschool through third" OR "preschool-3rd" OR "preschool-3"

The research team utilized the following terms for searches on the differentiated instruction topic:

("differentiat*" OR "individualiz*") AND (kindergarten* OR "grade 1" OR "first grade")

APPENDIX B. REFERENCE LIST FOR P–3 ALIGNMENT LITERATURE REVIEW

Appendix B.
Reference List for P–3 Alignment Literature Review

Quantitative Studies

Brown & Bogard (2007).

Reynolds, Magnuson, & Ou (2006).

Reynolds, Magnuson, & Ou (2010).

Mixed-Methods Study

Bogard (2006).

Qualitative Studies

Center for Applied Research and Educational Improvement (2013).

Center for the Study of Educational Policy (2012).

Jacobson, Jacobson, & Blank (2012).

Marietta (2010a).

Marietta (2010b).

Marietta & Marietta (2013a).

Marietta & Marietta (2013b).

Nyhan (2011).

Zellman & Kilburn (2011).

Theory and Policy Articles

“ABCs of early education: Listening, asking, sharing, engaging” (2013).

“10 action steps” (2011).

Advocates for Children of New Jersey (2010).

Bogard & Takanishi (2005).

Buenafe (2011).

Committee for Economic Development (2012).

Demanchick, Peabody, & Johnson (2009).

Donovan (2010).

Garland (2011).

Gates Foundation (2011).

Goldstein & Bauml (2012).

Grantmakers for Education (2006).

Groark, Mehaffie, McCall, & Greenberg (Eds.) (2007).

Guernsey, Bornfreund, McCann, & Williams (2014).

Halpern (2013).

Hernandez (2012).

Howard (2008).

Human Capital Research Collaborative (2014a).

Human Capital Research Collaborative (2014b).

Jacobson (2009).

Kagan, Carroll, Comer, & Scott-Little (2006).

Kagan & Kauerz (2010).

Kauerz (2006).

Kauerz (2009).

Kauerz & Coffman (2013).

King (2006).

Lesaux (2010).

Maeroff (2003).

Mead (2009).

NALEO Education Leadership Initiative (2008).

National Association of Elementary School Principals (2011).

New, Palsha, & Ritchie (2009).

The Pre-K Coalition (2011a).

The Pre-K Coalition (2011b).

Rice (2007).

Rice (2008a).

Rice (2008b).

Rice (2010).

Ritchie, Clifford, Malloy, Cobb, & Crawford (2010).

Ritchie, Maxwell, & Clifford (2007).

Ritchie, Maxwell, & Clifford (2009).

Scott-Little & Reid (2010).

Severns (2012).

Shore (2009).

Takanishi (2010).

Takanishi & Bogard (2007).

Takanishi & Kauerz (2008).

Tout, Halle, Daily, Albertson-Junkans, & Moodie (2013).

University of Chicago, Urban Education Institute, & Ounce of Prevention Fund (2012).

Appendix C. Reference List for Differentiated Instruction Literature Review

Appendix C. Reference List for Differentiated Instruction Literature Review

Quantitative Studies

Rigorous Designs (Analyzed for Potential to Meet WWC Design Standards)

Randomized Controlled Trials

Study that has the potential to meet WWC evidence standards without reservations.

Connor, Morrison, Fishman, Crowe, Otaiba, & Schatschneider (2013).[6]

Studies that do not appear to meet WWC evidence standards.

Al Otaiba, Connor, Folsom, Greulich, Meadows, & Li (2011).

Connor, Morrison, Fishman, Schatschneider, & Underwood (2007).

Connor, Morrison, Schatschneider, Toste, Lundblom, Crowe, & Fishman (2011).

Connor, Piasta, Fishman, Glasney, Schatschneider, Crowe … Morrison (2009).

Connor, Ponitz, Phillips, Travis, Glasney, & Morrison (2010).

Eastman (2010).

Quasi-Experimental Designs

Studies that do not appear to meet WWC evidence standards.

Arnold (2008).

Condron (2005).

Condron (2008).

Neel (2006).

Roth (2009).

Saylor (2008).

---

[6] This study presented multiple research designs. Only the within-grade design for first-grade effects has the potential to meet WWC evidence standards without reservations.

Non-Rigorous Designs

Descriptive

Hong, Corter, Hong, & Pelletier (2012).

Single-Group Pre-Test/Post-Test

Case-Smith, Holland, & Bishop (2011).

Geisler, Hessler, Gardner, & Lovelace (2009).

Menzies, Mahdavi, & Lewis (2008).

Qualitative Studies

Bofferding, Kemmerle, & Murata (2012).

Ensign (2012).

Holden (2007).

Kobelin (2009).

APPENDIX D. CODING PROTOCOLS

Appendix D. Coding Protocols

Coding Protocols for Qualitative Studies

Exhibit D1. Qualitative Study Protocol for PreK–3 Alignment

Topic of Interest	Instructions for Coder
Citation	Insert the citation of the article/study.
Citation #	Insert the internal Study ID number of the article/study.
Name of Program(s)	If applicable, insert the name of the program(s) that is the focus of the study or article.
Program Funding Source(s)	Insert any reference the author(s) makes to program funding source. This may include federal, state, local, or private funding sources. Also note in this section any information related to efforts to sustain PreK–3 alignment once funding is no longer available.
Resource Orientation	Insert the general reason for studying this topic, e.g., why the author(s) explains they decided to undertake the article/study.
Geographical Location	If provided, insert the name of the city(s) and state(s) where the study was conducted.
Setting	If provided, include where the authors collect their data (i.e., classroom, school, etc.)
Purpose of Study: Summary of Research Questions/Objectives	Provide the research questions and/or focal area of the study/article.
Sample Size and Participants	Include number of participants in study.
Qualitative Methods Used	Include methods authors used in study (i.e., interview, observation, etc.)
Definition of PreK–3 Alignment	Provide the authors' definition of PreK–3 alignment. For example, do they see this as PD between or among teachers from different grade levels? Do they define this as leadership? Is it curricula alignment?
Why PreK–3 Alignment is Important	If authors describe the potential benefits of PreK–3, include here.
Description of PreK–3 Alignment	Include the alignment of elements the author describes, e.g., alignment between: PD for preK–3 teachers, instruction and curricula; leadership and PD, etc.
Examples of PreK–3 Alignment	Examples of PreK–3 Programs: Include any examples of PreK–3 programs that author(s) references, e.g., Head Start, early childhood education centers, Chicago Parent Child Centers, Follow Through.
Key Elements of PreK–3 Alignment	Provide the core features the author(s) describes as key to the implementation of PreK–3 alignment. These may include common definitions; integrated family support services; structural features, etc.
Key Considerations for Implementing PreK–3 Alignment	Capture constructs author(s) identifies as key component for implementing PreK–3 alignment. Constructs may include leadership, joint PD, teacher quality, etc. and other elements that relate to curricula, instruction, ECE/school settings, and management/leadership.

Topic of Interest	Instructions for Coder
PreK–3 Alignment Challenges and Opportunities	PreK–3 Alignment Challenges: Provide a description of any obstacles or barriers to PreK–3 alignment that the author(s) discusses PreK–3 Alignment Opportunities: Provide the description of any opportunities or circumstances that lend themselves to PreK–3 alignment that the author(s) discusses.
Discussion of Outcomes	Include any outcomes that came about as a result of a program/ intervention in the article (e.g., increased reading proficiency for ELL students).
Summary of Findings/Conclusions	Provide key findings.
Describe Any Study Limitations (Noted by Author(s)	Include any study limitations the author mentions.
Reviewer's Comments on Study Limitations	Describe in the Annotation text box any problems or issues you note with the study or article. This may include, but is not limited to, the weakness of the study design, the quality of the methodology, etc.
Reviewer General Comments	This section is for any questions or comments reviewers have for discussion with project team and during interrater reliability.

Exhibit D2. Qualitative Study Protocol for Differentiated Instruction

Topic of Interest	Instructions for Coder
Citation	Insert the citation of the article/study.
Citation #	Insert the number of the article/study.
Name of Differentiated Instruction Program(s)	If indicated, insert the name of the program(s) that is the focus of the study or article.
Differentiated Instruction Program Funding Source(s)	Insert any reference the author(s) makes to the program funding source. This may include federal, state, local, or private funding sources. Also note in this section any information related to efforts to sustain Differentiated Instruction once funding is no longer available.
Resource Orientation	Insert the general reason for studying this topic, e.g., why the author(s) decided to undertake the article/study.
Geographical Location	If provided, insert the name of the city(s) and state(s) where the study was conducted.
Setting	If provided, include where the authors collect their data (e.g., classroom, school, etc.)
Purpose of Study: Summary of Research Questions/Objectives	Provide the research questions and/or focal area of the study/article.
Sample Size and Participants	Include number of participants in study.
Qualitative Methods Used	Include methods authors used in study (e.g., interview, observation, document review, etc.)
Definition of Differentiated Instruction	Provide the authors' definition of Differentiated Instruction. For example, do they see this as PD between or among teachers with students on different academic levels? Do they define this as classes broken down by levels across different school subjects (e.g., math, literacy)? Is it differentiation between classes (i.e., students divided up by levels in different classrooms) or differentiation within classes (differentiation that occurs at different times of the day, or by pairing students of different levels such that all can work on the same subject at the same time but using differentiated materials?)

Topic of Interest	Instructions for Coder
Why Differentiated Instruction is Important	If authors describe the potential benefits of Differentiated Instruction, include them here.
Description of Differentiated Instruction	Include the Differentiated Instruction the author describes, e.g., differentiated instruction for: math/literacy, between classrooms, within classrooms.
Examples of PreK–3 Differentiated Instruction	Include any examples of Differentiated Instruction programs that author(s) reference.
Key Elements of Differentiated Instruction	Provide the core features the author(s) describes as key to the implementation of Differentiated Instruction. These may include scaffolding behaviors, sequenced lessons, sequenced activities, letter and word study, quality curriculum, formative assessment, etc.
Key Considerations for Implementing Differentiated Instruction	Capture constructs author(s) identifies as key components for implementing Differentiated Instruction. Constructs may include leadership, PD, teacher quality, etc.
Differentiated Instruction Challenges and Opportunities	Provide a description of any obstacles/ barriers to Differentiated Instruction or circumstances that lend themselves to Differentiated Instruction that the author(s) discusses.
Discussion of Outcomes	Provide discussion of outcomes. Include any outcomes that came about as a result of a program/ intervention in article (e.g., increased reading proficiency for ELL students, academic outcomes, etc.)
Summary of Findings/Conclusions	Provide key findings.
Describe Any Study Limitations (Noted by Author(s)	Include any study limitations the author mentions.
Reviewer's Comments on Study Limitations	Describe any problems or issues you note with the study or article. This may include, but is not limited to, the weakness of the study design, the quality of the methodology, etc.
Reviewer General Comments	This section is for any questions or comments reviewers have for discussion with project team and during interrater reliability.

Coding Protocol for Theory and Policy Articles

Exhibit D3. Coding Protocol for Policy/Theory PreK–3 Alignment (coded in NVivo program)

Node/Definition/Coding Instructions	Sub-nodes/Definition/Coding Instructions
1. PreK–3 Alignment. Refers to a P–3 policy, program, or practice designed to improve U.S. children's early learning from preschool to third grade by aligning standards, curriculum, assessment, or professional development across these grades.	a. Examples of PreK–3 Programs: *Include any examples of PreK–3 programs that author(s) reference, e.g., Head Start, early childhood education centers, Chicago Parent Child Centers, Follow Through.*
	b. Setting of Program(s)*: Please include the setting of each program that is mentioned in the article, if applicable*
	c. Definition of PreK–3 Alignment: Provide the authors' definition of PreK–3 Alignment. For example, do they see this as PD between or among teachers from different grade levels? Do they define this as leadership? Is it curricula alignment?
	d. Key Elements of PreK–3 Alignment: *Provide the core features the author(s) describe as key to the implementation of PreK–3 alignment. These may include common definitions; integrated family support services; structural features, etc.* i. *Common definitions (Include examples of definitions the article provides on the PreK–3 topic, i.e., a definition for what PreK–3 alignment entails)* ii. *Integrated family support services(If any, include specific supports for students' families)* iii. *Structural features:* Include author's reference to aspects of PreK–3 program environment (e.g., the number of children per adult, the size of the class, the education and the training teacher, the presence or absence of a school-age program, the wages paid to teaching staff, teacher turnover rate, enrollment, etc.) iv. *Curricular Alignment Across Grades:* Provide ways in which the curriculum is aligned, such as using curricular materials (textbooks, programs) that are consistent from year to year v. *Preschool Onsite at Elementary School*: Preschool is in the same building as elementary school; considered a part of the school vi. *Full Day Kindergarten*: Kindergarten that has a morning and afternoon component and students attend both vii. *Consistent Learning Environment Across Grades*: Include any factors that help maintain consistency across grades (e.g., keeping small class sizes, using the same behavior management reinforcements and punishments across grade levels) viii. *Coordination Among Teachers*: Teachers communicate between grade levels to ensure that curriculum is well aligned from year to year and tailored to the right academic level based on incoming student data (i.e., if incoming Kindergarten class did poorly on letter-sound recognition in PreK, Kindergarten teacher takes this into account before jumping into more difficult concept) ix. *Small Class Size*: Code if article explicitly mentions small class size or says fewer than 20 students. x. *Gov't Leaders Support and Funds*: Provide and funding or support from Government Leaders, such as grants or political advocacy in favor of cause xi. *Smooth Transitions*: Provide information on how school promotes smooth transitions between grade levels (e.g., summer programs that prevent summer slide for at-risk students) xii. *Other*: Provide any other key elements

Node/Definition/Coding Instructions	Sub-nodes/Definition/Coding Instructions
	e. Key Considerations from Implementing PreK–3: *Capture constructs author(s) identify as key components for implementing PreK–3 Alignment. Constructs may include leadership, joint PD, teacher quality, etc.—elements that relate to curricula, instruction, ECE/school settings, and management/leadership.* i. *Leadership: Provide if authors mention leadership as a key component to successful PreK–3 alignment (e.g., strong principal leadership)* *Dedicating time for teacher, staff collaboration: Provide if authors mention dedicating time for teachers and staff collaboration as a leadership strategy* *Leader Support: Provide information on leadership support, if author provides* ii. *Joint PD: Provide if authors mention PD as a key consideration (e.g., PD on lesson planning techniques to ensure successful implementation of curriculum)* iii. *Comprehensive Early Childhood education: Provide mention of holistic early childhood education as a key consideration, e.g., developmentally appropriate practices* *Balance Academic and Developmental: Provide information on how academic and developmental needs are balanced (i.e., finding the right balance between play and academics, creating academic goals that are realistic and meet child's cognitive stage)* *Scale Up Proven Strategies: Provide, if relevant, how proven strategies are emphasized to maximize child's success* iv. *Partnerships with families: Provide information on family involvement as a key consideration, (e.g., consistent parent teacher communication about student progress* v. *Data sharing across ages: e.g., teachers sharing student data for student's new teacher to have a grasp of student's strengths and weaknesses* *1. Data focus: Provide, if relevant, how data is used and implemented* vi. *Teacher education & Degree requirements: Provide expected degree for role (e.g., teachers need a Master's in ELL education or equivalent work experience in order to be hired)* vii. *Funding Solutions: Provide sources of funding or funding strategies that enabled implementing PreK–3 program* viii. *Increase system cohesion: e.g., improved cohesion within school system, both hierarchical (i.e., communication between principal to teachers) and subject related (i.e., integrating subject areas to maximize student learning, such as reinforcing learning goals between subjects)* *Break Down Separate Systems: Provide, if relevant, how separate systems are broken down to create a more cohesive system (see viii. Description)* *Training on Alignment: Provide information, if relevant, on what training is made available to increase and improve alignment efforts* ix. *Cultural Responsiveness: Provide if/how school takes into account students' backgrounds, e.g., if parents do not speak English, have a translator on hand/ reports and other materials translated in order to promote home school communication*

Node/Definition/Coding Instructions	Sub-nodes/Definition/Coding Instructions
	f. PreK–3 Alignment Challenges: *Provide the description of any obstacles or barriers to PreK–3 alignment the author(s) discusses.* x. *Private PreK Wary of Public Schools: Provide, if relevant, Private PreKs being hesitant and/or uncooperative about merging with public schools* xi. *Leaders do not see PreK in purview: e.g., lack of efforts or incentive on behalf of political figures to push/advocate for Pre-K–3 alignment* xii. *Principal resistance: Provide any reason why Principal objects to PreK–3 alignment, if applicable* xiii. *Funding Barriers: Provide barriers to PreK–3 alignment related to/ caused by inadequate funding* xiv. *Knowledge of providers,: e.g., education or professional development providers have received that enable them to perform their role successfully*
	g. PreK–3 Alignment Opportunities: *Provide the description of any opportunities or circumstances that lend themselves to PreK–3 alignment that the author(s) discusses.*
Summary: This section summarizes that outcomes discussion and findings/conclusions of the study/article. It also captures whether the article may be described as advocating a particular policy or theory.	a. Advocacy Position or Statement: Indicate whether author states their position on the issue (does the author explicitly state/strongly suggest being in support or against a position or statement anywhere in the article?).

Coding Protocol for Quantitative Studies

Exhibit D4. Coding Protocol Studies Quantitative studies (adopted from WWC Study Review Guide for RCTs and Comparison Group QEDs [What Works Clearinghouse, 2010b])

Stage 1: Preliminary Screening for Descriptive Mapping Review

	Short Response	Supporting Information, Concerns, or Questions	Pages
Overview			
Intervention name: Name of the intervention(s) reviewed in this SRG. Note if 1 name for multiple versions or multiple names for 1 product			
Initial Screening			
Topic Area: Does the study focus on content that meets the definition for one of the three topics?	Yes/No		

	Short Response	Supporting Information, Concerns, or Questions	Pages
Focus: Is the intervention a program, product, policy, or practice as defined by the study's topic area?	Yes/No	Select a Focus: Program Product Policy Practice	
Time: Is the publication date in a target publication year?	Yes/No	Insert Publication Date	
Age or Grade Range: Does the study fit the age or grade range as specified in the review protocol?	Yes/No	Insert Age or Grade Range	
General Education: Does article fit the target sample as laid out in the study design?	Yes/No	Describe Sample	
Location: Does the study examine sample members in a location specified for the review protocol?	Yes/No	Insert State, Territory, or Tribal Area	
Outcomes: Does the study address at least one academic or cognitive outcome?	Yes/No	Describe Outcomes	
Screening Result: Does the study meet the screening criteria for the topic? Briefly explain if the study does not qualify.	Yes/No	If the study does not qualify, please provide a full explanation here	
Coding for Descriptive Mapping Review			
Design: What type of design is used to conduct the study (e.g., randomized controlled trial, quasi-experimental, regression discontinuity, single-case, case study, descriptive, correlational, theory, policy, ethnography, literature review, systematic review, meta-analysis, mixed methods, observational)? Select Yes in the Short Response column if the study used a randomized controlled trial or a quasi-experimental design, otherwise select No.	Yes/No	Select Design: Randomized trial Quasi-experiment Regression Discontinuity Single-case Case study Descriptive Correlational	
Sample Characteristics: Describe the sample characteristics of the study (e.g., gender, ethnicity, socioeconomic status)		Describe Sample Characteristics	
Effectiveness: Does the study examine the effect of an intervention?	Yes/No	Describe Intervention	
Study Comparison Group: Does the study use a comparison group?	Yes/No	Describe Comparison Condition	
Findings: Briefly describe the main findings reported in the study.		Describe Findings	
Screening for Evidence of Effectiveness Review			
Does the study meet the screening criteria for the effectiveness review? To meet the criteria the study must (1) use an RCT or QED design, (2) be an effectiveness study, and (3) use a comparison group?	Yes/No	If the study does not meet screening criteria for the effectiveness review, please provide a full explanation here	

Stage 2: Quality of Evidence for the Effectiveness Review (if the study passes Stage 1)

	Short Response	Supporting Information, Concerns, or Questions	Pages
Design Details			
How are the intervention and comparison groups formed?	Select Design: RCT Cluster RCT QED		
Is the study free of factors that are confounded with either group?	Yes/No		
Is there at least one relevant outcome that meets review requirements?	Yes/No		
Is there at least one outcome, sample, or time point with low attrition at the cluster and subcluster level?	Yes/No/NA		
Is evidence of baseline equivalence provided for at least one analytic sample, including statistical adjustment for characteristics relevant to equating the groups as given in the protocol, if needed?	Yes/No/NA		
Is the study free of other data or analytical issues that would affect the rating?	Yes/No		
What is the highest rating of an analysis in the study given current information? If more than one disposition code is appropriate, please copy and paste this row and select the additional disposition code(s).	Select Rating: Meets GDS without reservations Meets GDS with reservations Does not meet GDS	Select DNMGDS Disposition Code: The measures of effectiveness could not be attributed solely to the intervention The eligible outcomes did not meet WWC requirements Equivalence of the analytic intervention and comparison groups prior to the intervention was necessary and not demonstrated	
Explanation for Rating Disposition: If the study is rated Does Not Meet Group Design Standards, please provide a full explanation for the selected disposition code(s).			
If additional information is needed to complete the review, provide detail on the necessary information and how the rating could change			
If the rating may differ across study analyses, detail the rating for each sample, outcome, and time period combination, as necessary			

Stage 3: Study Details (if the study passes Stage 2)

	Short Response	Supporting Information, Concerns, or Questions	Pages
Did the authors present effect sizes? If so, how were they computed?	Yes/No		
Are estimates presented for subgroups in protocol?	Yes/No		
In summary, describe …			
Setting of the study (e.g., location, classrooms, courses, schools)			
Study design			
Sample sizes (e.g., students, classrooms, teachers, schools)			
Sample characteristics in protocol (e.g., race, gender, free/reduced lunch)			
Intervention condition as implemented in the study (including number of days/weeks/months, number of sessions, time per session)			
Comparison condition as implemented in the study			
Describe all eligible outcomes reported and how they were measured			
Are there outcomes that do not meet review requirements? If yes, provide the domain and a brief description of the reason why.	Yes/No		
Are there any outcomes that are not eligible for review? If yes, provide a brief description and the reason why.	Yes/No		
Support for implementation			

APPENDIX E. SUPPORTING DATA TABLES FOR RIGOROUS STUDIES ON DIFFERENTIATED INSTRUCTION

Appendix E. Supporting Data Tables for Rigorous Studies on Differentiated Instruction

Exhibit E1. Attrition, Baseline Characteristics and Findings for Connor, Morrison, Fishman, Schatschneider, and Underwood (2007) Study

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedge's g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Site-level sample size	NR	NR	NR	NR			NR	NR	
Student-level sample size	NR	NR	NR	NR			NR	NR	
WJ III Language and Literacy, Adjusted Mean Difference	NR	NR	NR	NR	-	2.63	NR	NR	NR

NR=not reported. WJ=Woodcock-Johnson Tests of Achievement.
NOTE: The randomized controlled trial (RCT) study had a total sample size at random assignment of 10 schools, 47 teachers and 616 first-graders. Authors did not present sample sizes by condition and did not discuss attrition. The authors did not report overall baseline means and standard deviations. They reported an adjusted mean difference with 95% CI = 0.37 to 4.90.
SOURCE: Connor, Morrison, Fishman, Schatschneider, and Underwood (2007).

Exhibit E2. Attrition, Baseline Characteristics and Findings for Connor, Piasta, Fishman, Glasney, Schatschneider, Crowe, and Morrison (2009) Study

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedge's g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Site-level sample size	5	5	5	5			5	5	
Student-level sample size	NR	NR	NR	NR			NR	NR	
Treatment condition x DFR interaction for amount of teacher/ child-managed, code-focused instruction on reading outcomes, HLM level-1 coefficient	NR	NR	NR	NR	-	-0.28	NR	NR	>.05
Treatment condition x DFR interaction for slope of teacher/ child-managed, code-focused instruction on reading outcomes, HLM level-1 coefficient	NR	NR	NR	NR	-	2.59	NR	NR	>.05
Treatment condition x DFR interaction for the amount of child-managed, meaning-focused instruction on reading outcomes, HLM level-1 coefficient	NR	NR	NR	NR	-	0.25	NR	NR	>.05

NR=not reported.
NOTE: This RCT study included a first-grade sample. The results in this study compared whether the intervention group individualized instruction closer to the A2i recommendations than the comparison group did. The study also compared reading growth in the intervention versus the comparison group while taking the distance from recommendation (DFR) into consideration. The DFR is the absolute value of the difference between the observed amount of time that a child receives a type of instruction and the amount of time that the A2i software recommends that a child should receive the type of instruction. The study presented means and standard deviations for fall and spring assessment data, but these data were not presented by assignment condition so they cannot be used for assessing overall treatment effects. The intervention group receives training and professional development on the A2i software and the comparison group does not. The outcome measures in this table are Woodcock Johnson standard scores; however, it is not clear from the published article whether these effects were for Letter Word Identification, Passage Comprehension, or Picture Vocabulary.
SOURCE: Connor et al.(2009).

Exhibit E3. Attrition, Baseline Characteristics and Findings for Connor, Ponitz, Phillips, Travis, Glasney, and Morrison (2010) Study

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Site-level sample size	5	5	5	5			5	5	
Student-level sample size	NR	NR	201	244			201	244	
WJ III Letter-Word Reading, Unadjusted Mean Difference	NR	NR	404.5 (28.04)	415.59 (32.47)	-0.36	-5.02	23.71	26.44	NR
WJ III Picture Vocabulary, Unadjusted Mean Difference	NR	NR	475.88 (10.39)	481.86 (10.47)	-0.57	-4.58	9.73	11.05	NR
WJ III Passage Comprehension, Unadjusted Mean Difference	NR	NR	447.35 (20.26)	451.76 (21.32)	-0.21	-3.28	15.15	15.7	NR
Head-Toes-Knees-Shoulder, Unadjusted Mean Difference	NR	NR	30.60 (8.96)	32.74 (5.99)	-0.29	-1.11	6.18	5.07	NR

NR=not reported. WJ=Woodcock-Johnson Tests of Achievement.
NOTE: The authors reported unadjusted means and standard deviations for the first-grade sample in this RCT study. Authors did not report individual p-values for mean differences. None of the mean differences for the findings were statistically significant. Random assignment occurred at the school level and the analysis used student-level data.
SOURCE: Connor, Ponitz, Phillips, Travis, Glasney, and Morrison (2010).

Exhibit E4. Self-regulation Findings for Connor, Ponitz, Phillips, Travis, Glasney, and Morrison (2010) Study, Hierarchical Linear Modeling (HLM) Results

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Site-level sample size	5	5	5	5			5	5	
Student-level sample size	NR	NR	201	244			201	244	
Head-Toes-Knees-Shoulders, HLM Adjusted Mean Difference	NR	NR	30.60 (8.96)	32.74 (5.99)	-0.29	-0.002	NR	NR	.247
Head-Toes-Knees-Shoulders, Fall Self-Regulation x A2i Use, HLM Adjusted Mean Difference	NR	NR	30.60 (8.96)	32.74 (5.99)	-0.29	-0.001	NR	NR	<.001

NR=not reported.
NOTE: In this RCT study with a first-grade sample, authors calculated mean difference using a level-2 HLM coefficient, where level-1 is the student level and level-2 is the classroom level (standard error of the coefficient=0.002). At the student level, the model controlled for fall test scores in Woodcock-Johnson (WJ) III Letter-Word subtest, WJ II Picture Vocabulary subtest, and Head-Toes-Knees-Shoulders. At the classroom level, the model controlled for percentage of students' low socioeconomic status (SES). The fall self-regulation x A2i software use interaction is a student level x classroom level interaction where fall self-regulation is a student level variable and A2i is a classroom level variable (standard error of the interaction coefficient=0.0002).
SOURCE: Connor, Ponitz, Phillips, Travis, Glasney, and Morrison (2010).

Exhibit E5. Attrition, Baseline Characteristics and Findings for Connor, Morrison, Schatschneider, Toste, Lundblom, Crowe, and Fishman (2011) Study

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Teacher-level sample size	NR	NR	NR	NR			NR	NR	
Student-level sample size	NR	NR	NR	NR			NR	NR	
WJ Letter-Word W score, Unadjusted mean difference	NR	NR	417.41 (29.64)	417.61 (32.01)	-	3.66	24.98	27.45	NR
WJ Letter-Word standard score, Unadjusted mean difference	NR	NR	107 (16)	108 (15)	-	0	14	14	NR
WJ Vocabulary W, Unadjusted mean difference	NR	NR	481.39 (9.23)	481.59 (14.87)	-	NR	NR	NR	NR
WJ Letter-Word Main Effect, HLM adjusted mean difference	NR	NR	NR	NR	-	7.84	NR	NR	.021
WJ Letter-Word Treatment x fall reading, HLM adjusted mean difference	NR	NR	NR	NR	-	-0.07	NR	NR	.236
WJ Letter-Word Treatment x fall vocabulary effect , HLM adjusted mean difference	NR	NR	NR	NR	-	-0.11	NR	NR	.550
WJ Letter-Word Treatment x special education status, HLM adjusted mean difference	NR	NR	NR	NR	-	-4.30	NR	NR	.575

Exhibit E5. Attrition, Baseline Characteristics and Findings for Connor, Morrison, Schatschneider, Toste, Lundblom, Crowe, and Fishman (2011) Study (Continued)

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
WJ Letter-Word Treatment x SES, HLM adjusted mean difference	NR	NR	NR	NR	-	-0.09	NR	NR	.234
WJ Letter-Word for students with lower fall reading (W=393), Cohen's d	NR	NR	NR	NR	-	0.59	NR	NR	NR
WJ Letter-Word for students with stronger fall reading (W=435), Cohen's d	NR	NR	NR	NR	-	0.41	NR	NR	NR
WJ Vocabulary for students with lower fall reading (W=474), Cohen's d	NR	NR	NR	NR	-	0.54	NR	NR	NR
WJ Vocabulary for students with stronger fall reading (W=487), Cohen's d	NR	NR	NR	NR	-	0.45	NR	NR	NR

NR=not reported. WJ=Woodcock-Johnson Tests of Achievement.
NOTE: Authors used a first-grade sample. All sample sizes are listed as not reported because the original study did not present sample sizes clearly.
SOURCE: Connor, Morrison, Schatschneider, Toste, Lundblom, Crowe, and Fishman (2011).

Exhibit E6. Attrition, Baseline Characteristics and Findings for Al Otaiba, Connor, Folsom, Greulich, Meadows, and Li (2011) Study

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
School-level sample size	7	7	NR	NR			NR	NR	
Student-level sample size	NR	NR	NR	NR			NR	NR	
WJ Letter Word standard score, unadjusted mean difference	NR	NR	95.53 (12.23)	97.27 (13.52)	-	-1.03	13.88	14.77	NR
WJ Word Attack standard score, unadjusted mean difference	NR	NR	96.37 (22.14)	98.86 (21.94)	-	-0.09	13.82	13.95	NR
AIMSweb Letter Sound Fluency, unadjusted mean difference	NR	NR	8.15 (9.61)	9.98 (10.26)	-	3.14	17.42	14.26	NR
DIBELS Phoneme Segmenting Fluency, unadjusted mean difference	NR	NR	NR	NR	-	12.13	22.97	15.61	NR
DIBELS Nonsense Word Fluency, unadjusted mean difference	NR	NR	NR	NR	-	2.51	24.66	23.04	NR

NR= not reported. WJ=Woodcock-Johnson Tests of Achievement. DIBELS=Dynamic Indicators of Basic Early Literacy Skills.
NOTE: In this RCT study with a kindergarten sample, the study authors did not report sample size information clearly enough to calculate attrition or establish baseline equivalence.
SOURCE: Al Otaiba, Connor, Folsom, Greulich, Meadows, and Li (2011).

Exhibit E7. Attrition, Baseline Characteristics and Findings for Al Otaiba, Connor, Folsom, Greulich, Meadows, and Li (2011) Study, Hierarchical Multivariate Linear Model (HMLM) Analysis

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
School-level sample size	7	7	7	7			7	7	
Student-level sample size	NR	NR	305	251			305	251	
WJ Letter Word z-score, HMLM adjusted mean difference	NR	NR	NR	NR	-	0.20	1.08	0.88	.022
WJ Word Attack z-score, HMLM adjusted mean difference	NR	NR	NR	NR	-	-0.02	0.98	1.03	.749
AIMSweb Letter Sound Fluency z-score, HMLM adjusted mean difference	NR	NR	NR	NR	-	0.05	0.99	1.01	.545

WJ=Woodcock-Johnson Tests of Achievement.
NOTE: In this RCT study with a kindergarten sample, the study authors did not report sample size information clearly enough to calculate attrition or establish baseline equivalence.
SOURCE: Al Otaiba, Connor, Folsom, Greulich, Meadows, and Li (2011).

Exhibit E8. Attrition, Baseline Characteristics and Findings for Al Otaiba, Connor, Folsom, Greulich, Meadows, and Li (2011) Study, Dynamic Indicators of Basic Early Literacy Skills (DIBELS). Hierarchical Multivariate Linear Model (HMLM) Analysis

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
School-level sample size	7	7	7	7			7	7	
Student-level sample size	NR	NR	303	245			303	245	
DIBELS Phoneme Segmenting Fluency z-score, HMLM adjusted mean difference	NR	NR	NR	NR	-	0.58	1.10	0.75	.000
DIBELS Nonsense Word Fluency z-score, HMLM adjusted mean difference	NR	NR	NR	NR	-	0.11	1.03	0.96	.223

NR=not reported.
NOTE: In this RCT study with a kindergarten sample, the study authors did not report sample size information clearly enough to calculate attrition or establish baseline equivalence.
SOURCE: Al Otaiba, Connor, Folsom, Greulich, Meadows, and Li (2011).

Exhibit E9. Attrition, Baseline Characteristics and Findings for Al Otaiba, Connor, Folsom, Greulich, Meadows, and Li (2011) Study, Latent Literacy HMLM Analysis

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
School-level sample size	7	7	7	7			7	7	
Student-level sample size	NR	NR	NR	NR			NR	NR	
Latent Literacy, HMLM coefficient	NR	NR	NR	NR	-	0.33	NR	NR	.002
Latent Literacy, HMLM adjusted model for Cohen's d using standard deviation =1	NR	NR	NR	NR	-	0.52	NR	NR	NR

NR=not reported. HMLM=Hierarchical Multivariate Linear Model.
NOTE: Authors randomly assigned seven schools to the intervention condition and seven schools were randomly assigned to the comparison condition. The study did not report enough information on the kindergarten sample to calculate attrition.
SOURCE: Al Otaiba, Connor, Folsom, Greulich, Meadows, and Li (2011).

Exhibit E10. Attrition, Baseline Characteristics and Findings for Connor, Morrison, Fishman, Crowe, Otaiba, and Schatschneider (2013) Study, Grade 1 Analysis

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Teacher-level sample size	NR	NR	NR	NR			NR	NR	
Student-level sample size	258	210	NR	NR			258	210	
WJ Letter-Word, Cohen's d from HLM adjusted mean differences, grade 1	258	210	NR	NR	-	0.32	NR	NR	.016
WJ Passage Comprehension, Cohen's d from HLM adjusted mean differences, grade 1	258	210	NR	NR	-	0.36	NR	NR	.016

NR=not reported. WJ=Woodcock-Johnson Tests of Achievement. HLM=Hierarchical Linear Model.
In this RCT study, authors randomly assigned teachers to treatment and comparison conditions. Both contrasts in this table have the potential to meet WWC Group Design Standards without reservations. The authors reported the assignment and analysis sample sizes as 28 teachers.
SOURCE: Connor, Morrison, Fishman, Crowe, Otaiba, and Schatschneider (2013).

Exhibit E11. Attrition, Baseline Characteristics and Findings for Connor, Morrison, Fishman, Crowe, Otaiba, and Schatschneider (2013) Study, Grade 2 Analysis

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Teacher-level sample size	NR	NR	NR	NR					
Student-level sample size	305	263	NR	NR			305	263	
WJ Letter-Word, Cohen's d from HLM adjusted mean differences, grade 2	305	263	NR	NR	-	0.44	305	263	.022
WJ Passage Comprehension, Cohen's d from HLM adjusted mean differences, grade 2	305	263	NR	NR	-	0.44	305	263	.022

NR=not reported. WJ=Woodcock-Johnson Tests of Achievement. HLM=Hierarchical Linear Model.
NOTE: In this RCT study, authors randomly assigned teachers to treatment and comparison conditions. Cohen's d is a standardized mean difference. The contrasts in this table did not report the information needed to test for baseline equivalence. The contrasts in this table have the potential to meet WWC Group Design Standards without reservations. The authors reported the assignment and analysis sample sizes as 49 teachers.
SOURCE: Connor, Morrison, Fishman, Crowe, Otaiba, and Schatschneider (2013).

Exhibit E12. Attrition, Baseline Characteristics and Findings for Connor, Morrison, Fishman, Crowe, Otaiba, and Schatschneider (2013) Study, Grade 3 Analysis

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Teacher-level sample size	NR	NR	NR	NR			NR	NR	
Student-level sample size	295	246	NR	NR			295	246	
WJ Letter-Word, Cohen's d from HLM adjusted mean differences, grade 3	295	246	NR	NR	-	0.25	295	246	.032
WJ Passage Comprehension, Cohen's d from HLM adjusted mean differences, grade 3	295	246	NR	NR	-	0.06	295	246	.032

NR=not reported. WJ=Woodcock-Johnson Tests of Achievement. HLM=Hierarchical Linear Model.
NOTE: In this RCT study, authors randomly assigned teachers to treatment and comparison conditions. Cohen's d is a standardized mean difference. The contrasts in this table did not report the information needed to test for baseline equivalence. The contrasts in this table have the potential to meet WWC Group Design Standards without reservations. The authors reported the assignment and analysis sample size as 40 teachers.
SOURCE: Connor, Morrison, Fishman, Crowe, Otaiba, and Schatschneider (2013).

Exhibit E13. Attrition, Baseline Characteristics and Findings for Connor, Morrison, Fishman, Crowe, Otaiba, and Schatschneider (2013) Study, Grades 1–3 Analysis

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Teacher-level sample size	NR	NR	NR	NR			NR	NR	
Student-level sample size	NR	NR	NR	NR			NR	NR	
Reading Factor Score, Cohen's d from cross-classified random effects growth-curve model, grades 1-3, comparing three years of treatment vs. three years of control	NR	NR	NR	NR		0.60	NR	NR	<.001

NR=not reported. WJ=Woodcock-Johnson Tests of Achievement. HLM=Hierarchical Linear Model.
NOTE: In this RCT study, authors randomly assigned teachers to treatment and comparison conditions. Cohen's d is a standardized mean difference. The Reading Factor Score, Cohen's d from cross-classified random effects growth-curve model, grades 1–3, comparing three years of treatment versus three years of control contrast was reviewed as a QED due to non-random placement of students into conditions. The contrasts in this table did not report the information needed to test for baseline equivalence. The authors reported a total analytic sample size of 95 teachers and 882 students.
SOURCE: Connor, Morrison, Fishman, Crowe, Otaiba, and Schatschneider (2013).

Exhibit E14. Attrition, Baseline Characteristics and Findings for Neel (2006) Study, Texas Primary Reading Inventory (TPRI) Analyses

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	80	86			83	85	
TPRI, Blend Words Task 1, Percentage	78.8% (40.87)	82.6% (49.88)	-	0%	-	-	-
TPRI, Blend Words Task 1, Percentage	60% (48.99)	46.5% (49.88)	-	4.6%	18.63	27.44	.206
TPRI, Detecting Initial Sounds, Percentage	42.5% (49.43)	15.1% (35.80)	-	5.7%	31.04	37.12	.289
TPRI, Detecting Initial Sounds, Percentage	42.5% (49.43)	15.1% (35.80)	-	5.7%	31.04	37.12	.289
TPRI, Detecting Initial Sounds, Percentage	42.5% (49.43)	15.1% (35.80)	-	5.7%	31.04	37.12	.289
TPRI, Detecting Final Sounds, Percentage	46.3% (49.86)	11.6% (32.02)	-	10.4%	25.85	38.08	.041
TPRI, Initial Consonant Substitution, Percentage	96.3% (18.88)	100% (0)	-	0%	-	-	-
TPRI, Final Consonant Substitution, Percentage	90.0% (30.00)	94.2% (23.37)	-	0%	-	-	-
TPRI, Middle Vowel Substitution, Percentage	78.8% (40.87%)	82.6% (37.91)	-	0%	-	-	-
TPRI, Initial Blending Substitution, Percentage	53.8% (49.86)	53.5% (49.88)	-	1.2%	10.89	15.30	.574
TPRI, Blends in Final Position, Percentage	50.0% (50.00)	37.2% (48.33)	-	1.1%	18.63	21.16	.723

NOTE: In this quasi-experimental (QED) study with a first-grade sample, the authors did not establish baseline equivalence on the analytic samples.
SOURCE: Neel (2006).

Exhibit E15. Attrition, Baseline Characteristics and Findings for Neel (2006) Study, Developmental Reading Assessment (DRA) Level Analysis

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	80	85			84	86	
DRA Level, Raw Score	6.69 (4.78)	5.19 (3.29)	-	-0.12	1.56	1.62	NR

NR=not reported.
NOTE: In this QED study with a first-grade sample, the authors did not establish baseline equivalence on the analytic samples.
SOURCE: Neel (2006).

Exhibit E16. Attrition, Baseline Characteristics and Findings for Neel (2006) Study, DRA Percent Analysis

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	79	85			83	73	
DRA, Percent	95.56 (3.63)	94.46 (3.08)	-	0.27	2.24	2.27	NR

NR= not reported. DRA=Developmental Reading Assessment.
NOTE: In this QED study with a first-grade sample, the authors did not establish baseline equivalence on the analytic samples.
SOURCE: Neel (2006).

Exhibit E17. Attrition, Baseline Characteristics and Findings for Neel (2006) Study, DRA Comprehension Analysis

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	79	85			83	72	
DRA, Comprehension	14.44 (5.10)	11.51 (2.57)	-	4.61	2.70	2.23	NR

NR=not reported. DRA=Developmental Reading Assessment.
NOTE: In this QED study with a first-grade sample, the authors did not establish baseline equivalence on the analytic samples.
SOURCE: Neel (2006).

Exhibit E18. Attrition, Baseline Characteristics and Findings for Neel (2006) Study, DRA Fluency Analysis

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	28	4			83	63	
DRA, Fluency	69.14% (24.84)	93.50% (39.57)	-	19.41%	31.23	22.69	NR

NR=not reported. DRA=Developmental Reading Assessment.
NOTE: In this QED study with a first-grade sample, the authors did not establish baseline equivalence on the analytic samples.
SOURCE: Neel (2006).

Exhibit E19. Baseline Characteristics and Findings for Saylor (2008) Study, Spring Post-Test Analyses

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Site-level sample size	3	3			3	3	
Student-level sample size	41	39			41	39	
GKAP-R Scores	64.63 (24.71)	52.56 (24.46)	-	4.34	19.53	20.98	NR
BLT, Scores	17.46 (7.60)	15.21 (6.28)	-	2.65	10.25	8.87	NR
DIBELS Initial Sound Fluency Scores	16.59 (8.72)	14.54 (10.31)	-	NR	NR	NR	NR
DIBELS Nonsense Word Fluency Scores	33.68 (14.82)	33.00 (14.74)	-	8.17	14.68	14.23	NR
DIBELS Phoneme Segmenting Fluency Scores	25.07 (19.50)	16.23 (15.53)	-	12.43	19.55	22.25	NR

NR=not reported. GKAP-R=Georgia Kindergarten Assessment Program – Revised. BLT=Basic Literacy Test. DIBELS=Dynamic Indicators of Basic Early Literacy Skills.
NOTE: In this QED study with a kindergarten sample, the authors used a prior-year cohort as the comparison group, which is not an acceptable QED design, based on WWC standards.
SOURCE: Saylor (2008).

Exhibit E20. Baseline Characteristics and Findings for Saylor (2008) Study, Two-Factor Analysis of Variance Change Scores

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Site-level sample size	3	3			3	3	
Student-level sample size	41	39			41	39	
GKAP-R, F-statistic	64.63 (24.71)	52.56 (24.46)	-	4.93	NR	NR	.03
BLT, F-statistic	17.46 (7.60)	15.21 (6.28)	-	0.09	NR	NR	.76
DIBELS Initial Sound Fluency, F-statistic	16.59 (8.72)	14.54 (10.31)	-	NR	NR	NR	NR
DIBELS Nonsense Word Fluency, F-statistic	33.68 (14.82)	33.00 (14.74)	-	10.58	NR	NR	<.01
DIBELS Phoneme Segmenting Fluency, F-statistic	25.07 (19.50)	16.23 (15.53)	-	1.05	NR	NR	.31

NR=not reported. GKAP-R=Georgia Kindergarten Assessment Program – Revised. BLT=Basic Literacy Test. DIBELS=Dynamic Indicators of Basic Early Literacy Skills.
NOTE: In this QED study with a kindergarten sample, the authors used an analysis of variance (ANOVA) F-test to test the winter to spring change score between treatment and comparison groups. This study used a prior-year cohort as the comparison group, which is not an acceptable QED design, based on WWC standards.
SOURCE: Saylor (2008).

Exhibit E21. Baseline Characteristics and Findings for Condron (2005) Study, Reading Gains for Grouped Versus Non-Grouped Peers

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	NR	NR			3374	1579	
IRT Reading Scale Score, Regression Coefficient	NR	NR	-	0.27	NR	NR	>.05

NR=not reported. IRT= Item Response Theory.
NOTE: In this QED study using data from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) data set, the author created an IRT Reading Scale Score, including measures of letter recognition, beginning sounds, ending sounds, sight comprehension of words, and comprehension of words in context. The author analyzed a sample of 668 schools.
SOURCE: Condron (2005).

Exhibit E22. Baseline Characteristics and Findings for Condron (2005) Study, Reading Gains for Low-, Middle-, or High-Skill Groups Versus Non-Grouped Peers

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	NR	NR			NR	NR	
IRT Reading Scale Score, Low-Skill Group vs. Non-Grouped, Regression Coefficient	NR	NR	-	-1.22	NR	NR	<.05
IRT Reading Scale Score, Middle-Skill Group vs. Non-Grouped, Regression Coefficient	NR	NR	-	0.76	NR	NR	>.05
IRT Reading Scale Score, High-Skill Group vs. Non-Grouped, Regression Coefficient	NR	NR	-	0.91	NR	NR	<.05

NR=not reported. IRT=Item Response Theory.
NOTE: In this QED using data from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) data set, the author created an IRT Reading Scale Score, including measures of letter recognition, beginning sounds, ending sounds, sight comprehension of words, and comprehension of words in context.
SOURCE: Condron (2005).

Exhibit E23. Baseline Characteristics and Findings for Condron (2008) Study, Low-Skill Groups Versus Non-Grouped Peers at First Grade

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	2219	4718			2219	4718	
IRT Reading Scale Score, Unadjusted Mean Difference	31.53 (8.57)	38.97 (13.11)	-	-14.42	15.39	20.11	<.001

IRT= Item Response Theory.
NOTE: In this QED study using data from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) data set, the author used imputed data for students with missing outcomes or pre-test scores. Baseline equivalence could not be tested on the analysis sample due to the author's use of imputed outcome and pre-test scores.
SOURCE: Condron (2008).

Exhibit E24. Baseline Characteristics and Findings for Condron (2008) Study, Middle-Skill Groups Versus Non-Grouped Peers at First Grade

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	3380	4718			3380	4718	
IRT Reading Scale Score, Unadjusted Mean Difference	36.45 (9.81)	38.97 (13.11)	-	-3.89	16.41	20.11	<.001

IRT=Item Response Theory.
NOTE: In this QED study using data from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) data set, the author used imputed data for students with missing outcomes or pre-test scores. Baseline equivalence could not be tested on the analysis sample due to the author's use of imputed outcome and pre-test scores.
SOURCE: Condron (2008).

Exhibit E25. Baseline Characteristics and Findings for Condron (2008) Study, High-Skill Group Versus Non-Grouped Peers at First Grade

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	3308	4718			3308	4718	
IRT Reading Scale Score, Unadjusted Mean Difference	46.44 (15.48)	38.97 (13.11)	-	11.74	19.42	20.11	<.001

IRT=Item Response Theory.
NOTE: In this QED study using data from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) data set, the author used imputed data for students with missing outcomes or pre-test scores. Baseline equivalence could not be tested on the analysis sample due to the author's use of imputed outcome and pre-test scores.
SOURCE: Condron (2008).

Exhibit E26. Baseline Characteristics and Findings for Condron (2008) Study, Low-Skill Groups Versus Non-Grouped Peers at Third Grade

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	1436	6873			1436	6873	
IRT Reading Scale Score, Unadjusted Mean Difference	58.42 (15.94)	70.95 (19.56)	-	-13.65	18.64	18.59	<.001

IRT=Item Response Theory.
NOTE: In this QED study using data from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) data set, the author used imputed data for students with missing outcomes or pre-test scores. Baseline equivalence could not be tested on the analysis sample due to the author's use of imputed outcome and pre-test scores.
SOURCE: Condron (2008).

Exhibit E27. Baseline Characteristics and Findings for Condron (2008) Study, Middle-Skill Groups Versus Non-Grouped Peers at Third Grade

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	2067	6873			2067	6873	
IRT Reading Scale Score, Unadjusted Mean Difference	64.82 (16.41)	70.95 (19.56)	-	-4.28	17.18	18.59	<.001

IRT= Item Response Theory.
NOTE: In this QED study using data from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) data set, the author used imputed data for students with missing outcomes or pre-test scores. Baseline equivalence could not be tested on the analysis sample due to the author's use of imputed outcome and pre-test scores.
SOURCE: Condron (2008).

Exhibit E28. Baseline Characteristics and Findings for Condron (2008) Study, High-Skill Groups Versus Non-Grouped Peers at Third Grade

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	2634	6873			2634	6873	
IRT Reading Scale Score, Unadjusted Mean Difference	79.43 (19.37)	70.95 (19.56)	-	7.62	15.59	18.59	<.001

IRT= Item Response Theory.
NOTE: In this QED study using data from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) data set, the author used imputed data for students with missing outcomes or pre-test scores. Baseline equivalence could not be tested on the analysis sample due to the author's use of imputed outcome and pre-test scores.
SOURCE: Condron (2008).

Exhibit E29. Attrition, Baseline Characteristics and Findings for Eastman (2010)

			Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention assignment sample**	**Comparison assignment sample**	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean differ-ence**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	27	27	23	22			23	22	
Total Number of Errors on Running Record Reading Assessment,[a] *Mean Difference*	27	27	6.48 (3.84)	6.68 (4.02)	-	0.55	3.11	3.48	0.583

[a] The outcome was not a standardized test and therefore does not have established reliability or validity; the authors did not provide additional evidence related to reliability and validity.
NOTE: In this RCT study with a first-grade sample, the author used a one-way analysis of variance (ANOVA; F=.306). Baseline equivalence could not be tested because the baseline measure does not have evidence of reliability or validity.
SOURCE: Eastman (2010).

Exhibit E30. Baseline Characteristics and Findings for Arnold (2008) Study

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Site-level sample size	5	16			5	16	
Student-level sample size	94	289			94	289	
TPRI Screening, Percent Growth	43% (49.51)	48% (49.96)	-0.12	1%	49.18	48.99	NR
TPRI Listening Comprehension, Percent Growth	45% (49.75)	56% (49.64)	-0.27	7%	40.00	33.63	NR

NR=not reported. TPRI=Texas Primary Reading Inventory.
NOTE: In this QED study with a kindergarten sample, the baseline data represent the percentage of proficient students whereas the effect size represents the difference between the percentage growth in the intervention group versus the percentage growth in the comparison group.
SOURCE: Arnold (2008).

Exhibit E31. Baseline Characteristics and Findings for Roth (2009) Study

	Baseline measure (standard deviation)	Baseline measure (standard deviation)		Findings	Findings	Findings	Findings
Variable	**Intervention group**	**Comparison group**	**Hedges' g**	**Mean difference**	**Intervention standard deviation**	**Comparison standard deviation**	**p-value**
Student-level sample size	48	53			48	53	
Writing Prompt Ideas, Gain Score	2.07 (0.59)	2.28 (0.74)	-0.31	0.55	0.64	0.66	<.001
Writing Prompt Organization, Gain Score	1.70 (0.75)	2.08 (1.01)	-0.42	0.74	0.81	0.87	<.001
Writing Prompt Word Choice, Gain Score	1.94 (0.50)	2.17 (0.64)	-0.40	0.50	0.61	0.68	<.001
Writing Prompt Sentence Fluency, Gain Score	1.90 (0.62)	2.20 (0.82)	-0.41	0.50	0.64	0.68	<.001
Writing Prompt Spelling, Gain Score	2.08 (0.69)	2.70 (0.85)	-0.79	0.91	0.64	0.59	<.001
Writing Prompt Spelling of High-Frequency Words, Gain Score	2.38 (0.79)	2.77 (0.78)	-0.49	0.68	0.82	0.76	<.001
Writing Prompt Capitalization, Gain Score	1.61 (0.45)	1.88 (0.62)	-0.49	0.34	0.72	0.64	<.05
Writing Prompt Punctuation, Gain Score	1.43 (0.48)	1.76 (0.51)	-0.66	0.29	0.74	0.74	<.05
Writing Prompt Spacing, Gain Score	1.36 (0.74)	1.42 (0.74)	-0.08	0.23	0.76	0.69	>.05
Writing Prompt Handwriting, Gain Score	2.40 (0.76)	2.35 (0.72)	0.07	0.58	0.85	0.74	<.001
Writing Prompt, Cohen's d	1.89 (0.47)	2.16 (0.59)	-0.50	1.3	0.42	0.49	<.0001
WJ Writing Sample, Cohen's d	6.83 (3.88)	10.95 (5.89)	-0.81	0.98	4.85	4.75	<.0001

WJ=Woodcock-Johnson Tests of Achievement.
NOTE: In this QED study with a first-grade sample, the author collected pre-test data to serve as baseline measures. In the findings columns, standard deviations reflect the gain score.
SOURCE: Roth (2009).

The Department of Education's mission is to promote student achievement and preparation for global competitiveness by fostering educational excellence and ensuring equal access.

www.ed.gov