Executive Summary

The purpose of this report is to present design options for a study of the effectiveness of different coaching dimensions in Head Start (HS) programs. This design project was funded by the U.S. Department of Health and Human Services (HHS), Administration for Children and Families (ACF), Office of Planning, Research and Evaluation (OPRE).

Under the task order, *Head Start Professional Development: Developing the Evidence for Best Practices in Coaching*, a design team was formed of four research organizations (American Institutes for Research [AIR], MDRC, MEF Associates, and Child Trends), which developed the design options presented here with input from consultants and practitioners in the HS field. The resulting study of coaching intends to:

- Provide strong evidence for effective and efficient coaching practices of center-based teachers of three- to five-year-olds in HS programs.
- Help HS programs make informed decisions about the allocation of professional development (PD) resources when designing, implementing, and improving coaching programs.
- Advance the state of empirical knowledge about coaching within typical early childhood settings and set the stage for additional future research about coaching as a professional development strategy.

The work of the design task order included (1) examining the conceptual and theoretical frameworks for coaching in early childhood education settings, (2) determining the best methodology for rigorously evaluating the effectiveness of coaching dimensions, and (3) designing a study (hereafter called the *HS Coaching Study*) to evaluate specific dimensions of coaching that may impact teacher and classroom practices in HS and other early childhood settings. A dimension refers to a singular aspect or component of a coaching program (e.g., coach characteristics, type of coaching activity, dosage); the study will examine the effect of *varying the levels* of coaching dimensions.

This report provides recommendations for the following aspects of the HS Coaching Study:

- The purpose of the study
- The research questions
- The study design for testing the impact of coaching, including the following:
    - Application of the multiphase optimization strategy (MOST) framework
    - Systematic evaluation of three dimensions of coaching (dosage of coaching, recipient of coaching, and amount of coach training)
    - Use of a factorial design
    - Requirements for detecting effects and sample size
- The implementation research component of the study
- The cost component of the study
- The measures

- The important logistical issues for this study, such as participant recruitment, participant selection, the implementation monitoring, and the technical assistance that may be required

The report also provides information about the content of the coaching intervention and the standardized foundational coaching approach for the study. Although some approaches to coaching do not specify a particular content domain on which teachers and coaches will concentrate, we suggest that the goals of this study will be better met, and outcomes more precisely measured, by using a coaching approach with a specific content focus. After considering a number of content areas geared towards supporting various domains of early childhood development, we recommended that the HS Coaching Study focus on language development and the interactions between children and teachers that support that development. Language development is a critical domain of early child development, a well-established precursor to subsequent literacy skills that grow increasingly important as children approach entry to elementary school. It is one of the 11 domains within the HS Child Development and Early Learning Framework.

Descriptions of the process, criteria, and guiding principles are used throughout the report to support the design recommendations for the study. To help in planning for the HS Coaching Study, we provide estimates of the resources needed to conduct this study, suggested task structure, and a study timeline.

The Purpose of the HS Coaching Study

There is a growing consensus in the early childhood education (ECE) field that the provision of targeted high-quality professional development shows promise for improving teachers' practices, classroom quality, and child outcomes (Diamond & Powell, 2011; Dickinson & McCabe, 2001; Snyder, Hemmeter, & McLaughlin, 2011). Coaching is a recommended PD practice that is increasingly widespread. Although the available evidence generally supports the positive effects of coaching overall, there are significant challenges with interpreting the evidence for the effectiveness of coaching components.

Two extant literature reviews on coaching in ECE (Aikens & Akers, 2011 and Isner et al., 2011) noted a number of limitations. Most importantly, many studies did not provide detailed specifications about the coaching in their interventions. Overall, key limitations to extant coaching research are:

- Coaching is usually examined in combination with additional PD strategies; coaching is part of effective PD packages and is seldom studied on its own.
- Descriptions of coaching features (e.g., structure, process, and staffing aspects of coaching programs) lack sufficient detail.
- The most effective coaching actions and behaviors have not been identified through experimental methods. Coaching features are not examined separately in the extant literature. Little empirical support has been presented for the value from adding certain coaching strategies as part of a PD program (e.g., adding training for coaches).

- Few coaching studies have systematically examined the effectiveness of variations of coaching dimensions (e.g., how much training for coaches is most effective?).

There is a traditional PD paradigm for many evaluations—testing whole interventions rather than individual dimensions. In most evaluations of coaching, coaching *content* may be bundled, or combined, with delivery in a particular *format*, bundled with a particular *dosage* of the intervention, which is further bundled with delivery to a particular *recipient*. This combination of coaching features may then be combined with additional curriculum training and materials provided to teachers in a PD package. However, it leaves evaluators, policymakers, and program developers with an intervention "black box," for which it is hard to understand which individual dimensions influence outcomes.

The design for the HS Coaching Study aims to strengthen the research by evaluating coaching, as a stand-alone professional development component in the HS context and to examine the differential effects of several specific dimensions of coaching.

The Guiding Research Questions

Six research questions guided the design of the HS Coaching Study, two related to the impact of coaching dimensions, three related to the implementation of coaching dimensions, and one related to cost.

The key questions related to *impact* of the coaching dimensions are:

1. What is the effect of specific dimensions of coaching on teacher practices and classroom quality in HS programs?
2. Does the effect of one coaching dimension depend on the level of another coaching dimension?

The research questions related to the *implementation* of the coaching dimensions are:

3. Are the different coaching dimensions implemented with fidelity?[1]
4. What factors facilitate or challenge the fidelity of implementation of the different coaching variations?
5. How does implementation vary across grantees' program environments, populations, and other contextual program features?

The research question related to *cost* is:

6. What is the cost of implementing the different coaching dimension variations?

[1] Fidelity here refers to implementation of the coaching dimensions as designed. There may be several aspects of fidelity that may be of interest, including adherence, exposure, responsiveness, and quality.

Impact Component of the HS Coaching Study

The MOST Approach

As part of the task order, the design team wrote a review that outlined different possible design and methodology framework options for the study (Somers, Collins, Maier, 2013; http://www.acf.hhs.gov/programs/opre/research/project/head-start-coaching-study-design-phase). After reviewing a range of research methods for testing the effectiveness of coaching, the design team and OPRE staff members decided that the design for the HS Coaching Study should reflect the principles of the multiphase optimization strategy (MOST; Collins et al., 2005; 2009; in press).

The MOST framework is a staged and rigorous approach to developing and evaluating interventions.

- After a preparation phrase, an Optimization Phase is conducted, in which the relative effect of different intervention dimensions are assessed in a randomized screening experiment. Dimensions are selected for testing by examining the evidence base or, if the evidence base is weak, using strong theoretical support or recommendations from experienced practitioners and researchers.
- The results of this screening experiment are then used to build an optimal intervention model consisting of the selected dimensions that meet some minimum threshold for effect size, cost-effectiveness, and practical or theoretical importance.

In a second phase, the impact of this optimal model is evaluated in a standard two-group randomized experiment. The HS Coaching Study corresponds to the Optimization Phase of the MOST framework.

Systematic Evaluation of Coaching Dimensions

With the MOST approach as a guiding framework, we recommend that the HS Coaching Study examine the effect of three individual coaching dimensions:

(1) The amount or dosage of coaching (Dosage);

(2) The recipient of the coaching (Recipient; lead teacher only vs. teaching team); and

(3) The amount of coach training (Coach Training) or Delivery Mode (Mode; technologically-mediated vs. onsite)

Strictly speaking, we recommend that the study examine the effect of *varying the levels* of each of these coaching dimensions. For example, for Dosage, we suggest testing outcomes of having coaches meet with teachers on a bi-weekly vs. monthly basis.

Factorial Design for the Impact Study

To examine the dimensions, we suggest that a factorial design is the most suitable design for testing the effect of the three coaching dimensions. A factorial design is an experimental design in which the experimental conditions represent all possible combinations of the levels of the dimensions under investigation. Factorial experiments are well suited for building strong interventions in the Optimization Phase of the MOST framework (e.g., Collins et al., 2005; in press). Specifically, for three coaching dimensions, we recommend a factorial design with three factors and eight experimental conditions, as the table below illustrates.

Recommended 2^3 Factorial Design

Experimental Condition Number	Factors		
	Amount of Coaching (*DOSAGE*)	**Recipient of the Coaching (RECIPIENT)**	**Amount of Coach Training (TRAINING)***
1	Monthly	Lead teacher only	Orientation
2	Monthly	Lead teacher only	Ongoing
3	Monthly	Teaching team	Orientation
4	Monthly	Teaching team	Ongoing
5	Biweekly	Lead teacher only	Orientation
6	Biweekly	Lead teacher only	Ongoing
7	Biweekly	Teaching team	Orientation
8	Biweekly	Teaching team	Ongoing

Note. Unshaded cells represent the typical level (Level I) of the factor; shading denotes the enhanced level (Level II) of the factor.
*Or Mode, in which case the levels in the design would be remote coaching (in the unshaded cells) and in-person coaching (in the shaded cells).

Although factorial designs require more experimental conditions than other designs, a benefit is that they require a smaller sample size than other designs to statistically detect a dimension's effect of given magnitude. Another potential benefit of factorial designs is that they also account for—and provide information on—interaction effects between the dimensions that are being tested in the study. Thus, factorial designs make it possible to efficiently determine which particular components of an intervention are more important, as well as examine how these components interact with each other to produce the desired outcomes. For these reasons, factorial experiments provide findings that are useful for policymakers and practitioners who are creating or adapting interventions.

Minimum Detectable Effect Size and Sample Size

The report provides a full explanation of the power and sample plan for the HS Coaching study. The *minimum detectable effect size* (MDES) is a useful concept for making decisions about the sample size. Formally, MDES is the smallest true effect on the outcome of interest (scaled as an effect size) that can be detected with a reasonable degree of power. The recommendation is that the HS Coaching Study be able to detect a main effect on teacher and classroom outcomes of

0.20. The recommendation has two justifications. First, it seems reasonable to expect that the coaching dimensions in the study would have main effects of this size on teacher practices. Based on prior research, an additional 1.5 hours of coaching per month (which is the one of the variations that will be tested in the HS Coaching Study) could improve teacher practices by an effect size of about 0.09 to 0.26, with effects expected to larger for practices that teachers used less frequently at baseline.[2] Thus, it is reasonable to expect that the dimensions under study could have a main effect of 0.20 on teacher practices that are in greatest need of improvement. Second, it is probable that an effect size of 0.20 on teacher practices can also translate into a meaningful change in children's literacy-related outcomes. Even though child outcomes will not be measured in the HS Coaching Study, improving children's outcomes is one of the goals of coaching. An effect size of 0.20 on teacher and classroom outcomes translates into an effect of approximately 1.4 to 2.5 weeks of extra learning for children, or a 5 to 10 percent increase in children's literacy skills above and beyond what they would normally learn during the school year.[3]

We estimate that in the proposed factorial design approximately 248 centers across 31 HS grantees will be needed to detect an effect size of .20 if random assignment occurs at the center level. However, the final sample size will depend on (a) final decisions that OPRE and the study evaluation team make about the specifications of random assignment (whether dimensions are assigned at the coach level, and/or whether a HS grantee would allow the evaluators to randomly assign coaches to centers for the purposes of the study) and (b) how many classes exist per center for participating sites.

Implementation Component of the HS Coaching Study

Implementation research helps document the extent to which the intervention was implemented as intended. Implementation research identifies factors that may facilitate and challenge execution of the intervention that further contextualize the resulting impacts. For the HS Coaching Study, we recommend the following goals:

(1) To describe and assess the fidelity of implementation for the eight experimental coaching conditions in order to help interpret impacts.

(2) To inform future development of effective and feasible coaching models.

Documenting the foundational coaching model (including the implementation of the language content of the coaching) and the three systematically varied dimensions will be important to understand fidelity (i.e., the extent to which the coaches and teachers implement the levels of the targeted three dimensions—Dosage, Recipient, and Coaching Training—to which they were assigned) and the extent to which coaches and teachers adhere to the dimensions that are fixed and the natural variation across the teachers and coaches for other dimensions.

[2] This is based on a study conducted by Landry and her colleagues (2009), which found that four additional hours of coaching per month can improve teacher practices by an effect size of 0.23 to 0.70.
[3] Estimates of annual effect size gains are based on data from the CARES study (Mattera, Lloyd, Fishman, & Bangser, 2013).

Cost Component of the HS Coaching Study

If the evaluation team learns that particular coaching dimensions are effective, the total resources required to implement these dimensions will be important information for both planners within OHS and HS program directors. The cost aspect of the HS Coaching Study aims to accomplish the following goals:

(1) Provide information to HS grantees about the types of resources needed to develop and implement the targeted coaching dimensions within their programs.

(2) Gather information that can be used in a cost-effectiveness analysis.

Conducting this analysis would allow the evaluation team to determine the relative cost-effectiveness of each coaching dimension condition by comparing the financial resources required to implement a given level of a coaching dimension (e.g., low dosage of coaching or enhanced coach training) and its estimated effectiveness (effect size) when considered across all other dimension levels.

Measurement

The measurement approach, was designed to maximize study feasibility (conducting the study within the timeline and minimizing burden on participants) while simultaneously documenting the details and context of coaching with the necessary richness and specificity to answer the research questions.

Key constructs for the study were identified based on the research questions. We then provide details for specific recommended data collection tools for the impact, implementation, and cost research, including what they measure and their format, frequency, and specifications. Most suggested data collection tools serve multiple purposes in the HS Coaching Study. The measurement strategy is not simple. However, it is important to collect data with multiple respondents and at multiple levels to understand the complex practices that are part of the HS Coaching Study. We suggest six categories of data collection tools in addition to requesting program budgets. These are listed below:

(1) Implementation Contact, Time, and Attendance Logs
Participants: Coaches, PD trainers, teachers (using time sampling)
Purpose: Document and monitor attendance and details of coaching sessions, coach training, and teacher training

(2) Implementation Rating Logs
Participants: Coaches, PD trainers
Purpose: Document (a) coaches' report on utility and value of coach training; (b) coaches' and teachers' reports on utility and value of teacher training; (c) coaches' and teachers' reports on utility and value of coaching sessions; and (d) coaches' and trainers' reports of teachers using targeted strategies

(3) Participant Surveys
Participants: Center directors, coaches, teachers, PD trainers

Purpose: Gather data about participant characteristics, experiences, and perceptions of coaching

(4) Participant Interviews
Participants: Center directors, grantee liaisons; sample of coaches, teachers, PD trainers
Purpose: Gather data about how coaching was implemented, factors that facilitated or hindered implementation and fidelity

(5) Observations of Coaching Sessions and Coach Training
Participants: Coaches, teachers
Purpose: Assess key qualitative features of the coaching sessions

(6) Observations of Teacher Practices and Classroom Environment
Participants: Teachers
Purpose: Gather impact data about (a) classroom quality and (b) the specific language and teacher-child interaction practices that are targeted by the coaching

Conducting the Study

Relevance for the field

Aiming to design a study that is as relevant and compelling as possible for the HS field, as well as logistically feasible, as part of the design process we consulted with a limited number of stakeholders at OHS and in the HS practitioners. We spoke to stakeholders about either their experience with coaching programs or their opinion about coaching in general or in the context of the planned study. Feedback was gathered through individual calls, group webinar-format calls, and at an interactive conference presentation.

Logistical Issues

Problems related to implementation of the foundational coaching model and the eight coaching conditions could inevitably arise in a complex study in up to 31 grantees and 248 centers. Therefore, we recommend carefully explaining the study to potential participants, monitoring implementation, and providing assistance as necessary. The logistical issues examined by the coaching team include:

- Recommendations for participant recruitment and selection, including establishing partnership with OHS and HS grantees and consideration of funding the coaching efforts for each participating grantee.

- Monitoring of implementation and technical assistance, including establishing clear expectations, assigning an onsite liaison with each participating grantee to facilitate communication; and structured technical assistance. .

Conclusion

Using the first phase of the MOST framework to guide the HS Coaching Study design will allow for the systematic testing of the impact of coaching dimensions when controlling for all other variations studied. Certainly, adapting the MOST model to the complexities of the HS coaching interventions is not easy. However, the answers to the research questions for the HS Coaching

Study related to coaching impact, implementation, and costs will play an important role in informing HS programs decisions about the allocation of their PD resources when developing and implementing coaching approaches. In addition, the answers to these proposed research questions will advance the research evidence about coaching in early childhood settings. Ideally, results from the HS Coaching Study will help in designing an optimal coaching intervention that will be the focus of additional research.