

## Executive Summary

This report is part of a larger effort to design a study to evaluate the effect of individual coaching components in Head Start programs. The design project is guided by the following research question: *What is the effect of individual coaching components on teachers and children in the Head Start context?* The goal is to design an evaluation that will help Head Start programs, and other early childhood programs, implement stronger coaching interventions by providing them with reliable evidence on the effect of coaching components that they can use to decide which components to implement, given their local needs and budgetary constraints. The purpose of this present report is to review different experimental designs that could be used to estimate the effect of individual components within a social intervention, such as Head Start coaching.

In the research design literature, an intervention *component* is defined as any aspect, element, or feature of an intervention that can be reasonably separated out in order to study its individual effect on the outcomes of interest. For example, in Head Start and other early childhood education settings, coaching interventions consist of multiple components that are intended to improve teacher practice and classroom quality, and ultimately child outcomes. A coaching intervention may include program components related to structure or delivery (e.g., coach credentials, coach training, coach caseload, coach supervision) and components related to the content or process of coaching (e.g., use of modeling; quantity and nature of feedback to the teacher). Each component has possible values, or *levels*. A component may be “on” or “off” in an intervention, or it can take on varying levels of intensity (e.g., “low” versus “high”).

Unfortunately, there is little rigorous evidence on the effect of individual intervention components. For this reason, decisions about which components to include in a social intervention such as coaching are based primarily on theory and professional experience about which components are likely to matter, rather than empirical evidence on the effect of these components. This means that social interventions may not be as effective or as cost-effective as they could be. If the effect of individual components were known *a priori*, this information could be used to design interventions that are not only more effective but also less time consuming and more economical. In order to build interventions that have maximum impact, and that are flexible to local context and needs, evaluation science needs to move towards policy experiments that test the effect of individual intervention components.

Accordingly, the goal of this report is to review potential experimental design options that could be used to estimate the effect of individual coaching components in Head Start. Five experimental designs are discussed: factorial designs, comparative treatment designs, the individual experiments design, crossover designs, and adaptive clinical trials. The differences between these designs are elucidated in terms of how well they can answer the study’s research question; their sample size requirements; the number of experimental conditions that would have to be implemented; and whether interactions between components can be estimated.

The main conclusion from this review is that a factorial design is the strongest experimental design for evaluating the effect of individual intervention components, such as coaching components in the Head Start context. Factorial designs provide findings that are useful for policymakers and practitioners who are creating or adapting interventions in the field because they account for—and provide information on—interaction effects between components. Although evaluators often disregard factorial designs because they require many more experimental conditions than other designs, they also require a smaller sample size to statistically detect a component effect of a given magnitude. This can outweigh the disadvantage and cost of having to implement a larger number of conditions. The other four designs reviewed in this report are more suitable for different purposes—namely comparing the effect of different intervention models (as opposed to components) or estimating the effect of a single component.

The report concludes by describing several issues that need to be considered when designing a study of component effects, regardless of which experimental design is used. A unique challenge with studies of component effects is that the expected effect of a single component is likely to be smaller in magnitude than the effect of an entire intervention. This means that the total sample size needed for a study of component effects will likely be larger than the sample size needed for an evaluation of a complete intervention, and therefore it becomes especially important to use strategies to improve statistical power (e.g., use of baseline covariates, choosing a lower level of random assignment, using well-aligned and reliable outcome measures, etc.). In a study of component effects, evaluators must also decide whether or not to “fix” the levels of non-tested components, and they must gauge the study’s feasibility in the field, because a test of multiple components is more operationally complex than an evaluation of a single intervention.