Abstract Title Page
Not included in page count.

Title: Methods for Investigating the Role of Program Quality in Determining Head Start’s Impact on Child Development

Authors and Affiliations:
Laura R. Peck, Ph.D., Principal Scientist
Abt Associates Inc., Social and Economic Policy Division
4550 Montgomery Ave., Suite 800N, Bethesda, MD 02814-3343
301-347-5537 / laura_peck@abtassoc.com

Stephen H. Bell, Ph.D., Principal Scientist, Senior Fellow and Vice President
Abt Associates Inc., Social and Economic Policy Division
4550 Montgomery Ave., Suite 800N, Bethesda, MD 02814-3343
301-634-1712 / stephen_bell@abtassoc.com
Abstract Body
Limit 4 pages single-spaced.

Background / Context:
Description of prior research and its intellectual context.
The Head Start Impact Study (HSIS) has shown that having access to Head Start improves children’s preschool experiences and school readiness on average across centers of high and lower quality, with some advantages persisting through first grade (U.S. Department of Health and Human Services, 2010). That said, scholars and practitioners alike have wondered whether impacts might be larger or more persistent for those who participate in particularly high quality Head Start centers. This study contributes to filling this substantive gap by making methodological innovations.

Purpose / Objective / Research Question / Focus of Study:
Description of the focus of the research.
Our research undertakes an analysis of Head Start quality that capitalizes on the experimental evaluation design. In this paper we explain the analytic process for measuring the role of quality in determining program impact and how the method (described below) can be applied to the Head Start data. To the extent that findings are available by Spring 2013, we will also report how Head Start quality influences effects on children’s short-term development in the cognitive, health, behavioral and social spheres.

Setting & Subjects:
Description of the research location.
(May not be applicable for Methods submissions)
The HSIS includes a nationally representative sample of 4,667 three- and four-year-old children in 383 centers nationwide. The HSIS has followed these children through their third grade year, with rich data available across many outcomes over multiple years. Important to this research is the observational data that describes program quality within Head Start centers.

Intervention / Program / Practice:
Description of the intervention, program, or practice, including details of administration and duration.
(May not be applicable for Methods submissions)
Head Start provides developmentally-focused pre-school education to children from disadvantaged families with the aim of ensuring their Kindergarten readiness in cognitive, behavioral and social domains.

Significance / Novelty of study:
Description of what is missing in previous work and the contribution the study makes.
The challenge that analyzing the role of quality poses is that children who participate in high quality Head Start are likely to differ from those who participate in lower quality Head Start in important ways that relate to their outcomes, independently of the nature of their Head Start experiences. This issue of selection arises in analyses of quality or dosage in K-12 schooling or for any educational intervention studied with a randomized experimental design, making the proposed paper of keen importance to the field of program evaluation most narrowly and to the many other fields of study related to education more broadly. The approach, while established, is new to application in the Head Start setting. Its extension here has provided opportunities to
consider other extensions of the work to inform important “what works” questions in many areas of social policy.

**Statistical, Measurement, or Econometric Model:**

*Description of the proposed new methods or novel applications of existing methods.*

As noted above, comparing those children in high quality Head Start with those in low quality Head Start or with no Head Start exposure—all within the study’s treatment group only—would result in impact estimates biased by selection. These samples would differ at baseline on unmeasured characteristics that lead to different outcomes independently of the effects of different quality Head Start experiences. To avoid this problem and capitalize on the experimental design of the HSIS, this work extends the approach established in Peck (2003) to the Head Start case. The technique identifies predicted high quality sample members from the treatment group and the control group in identical fashion, then estimates impacts on that subpopulation as one would in any experimental subgroup analysis. The symmetry of the identification procedure ensures that equivalent subgroups are compared and guarantees that the impact estimates are free from differential selection bias or other sources of internal bias.

While symmetric selection of treatment and control subgroup members within the experimental data ensures unbiasedness of the impact estimates generated for the subgroups examined, the subgroup for which the methodology produces unbiased impact estimates—children with the highest predicted probabilities of being in high quality Head Start programs, for example—is not necessarily the subgroup of policy interest—children who actually experience high quality Head Start. The predictive model, while symmetric for both treatment and control groups, is imperfect for both groups, potentially reducing the relevance (i.e., the external validity or generalizability) of the findings. This is why we apply procedures to convert results so that they represent impacts on actual rather than predicted subgroup members, subject to certain assumptions. The following steps are involved in carrying out the research, each of which is elaborated below:

1. Select a random subsample of the treatment group from which to predict the level of Head Start quality.
2. Using baseline characteristics, predict quality.
3. Use the resulting predicted quality variable to identify subgroups symmetrically in the treatment and control groups.
4. Analyze the impact of predicted quality by comparing mean outcomes between the symmetric treatment and control group subgroups created.
5. Convert results for predicted subgroups to represent impacts on actual subgroups under certain assumptions.

**Step 1. Select a random subsample of the treatment group to predict Head Start Quality.**

A key feature of this approach to subgroup analysis is retaining the strength of the experimental design. In order to do this, an important first step is to create a modeling subsample to use in identifying symmetric subgroups within the treatment and control groups, subgroups with equivalent predicted probabilities of participating in Head Start at a particular level of quality. Using the entire treatment group for subgroup prediction and for impact analysis could introduce bias because of the better fit that is inevitable for the sample that is used for modeling. To clarify, if the whole treatment group were used for prediction, then the model might more accurately identify the desired subgroup for treatment group cases than for predicted control group cases. This is because the prediction model would mold its parameters to the errors that exist in the outcome data due to random baseline variation between the groups. This would
result in some unknown amount and direction of bias that is easily avoidable by designating separate predictive and impact estimation subsample of the treatment group. In this application, we will select a random half of the treatment group for predictive modeling, retaining the other half for impact estimation.

**Step 2. Using baseline characteristics, predict quality.** In our application, we will create three distinct quality indicators for all members of the age 3 and age 4 treatment groups, each with three levels: a value of 0 representing those who never participated in Head Start; a value of 1 representing “low quality” Head Start, among those who participated in the program; and a value of 2 representing “high quality” Head Start, also among those who participated in the program. The specific threshold for dividing high quality from low quality is measure-specific as detailed in the paper. With this categorical quality measure as our dependent variable, we used a generalized logit procedure to predict quality, with explanatory variables including center, family, and child characteristics as follows:

- **Center Characteristics:** center of random assignment (series of dummy variables, omitting the dummy for one center)
- **Family Characteristics:** home language, both bio-parents at home, primary caregiver’s age, mother’s education, bio-mother’s recent immigrant status, mother’s marital status, mother gave birth to study child as a teen
- **Child Characteristics:** sex, age, race, language

We expected that the center of random assignment will be the best predictor of the quality of Head Start; we further allow this to proxy other community characteristics that might be associated with higher quality. Other family- and child-level characteristics might also be associated with the quality of Head Start that a child experiences. Rather than basing our decision for which predictor variables to include on arbitrary or theoretical factors, we follow the lead of propensity score methods (to which our treatment group predictive modeling procedure is closely akin) which advocate a “kitchen sink” approach for generating the greatest explanatory power and best correct prediction rate possible. We are uninterested in interpreting any of the coefficients on our explanatory variables from the prediction model but instead have as our goal the best “hit rate”: correctly matching those predicted to be in each of our three subgroups with their actual subgroup experience.

**Step 3. Use resulting predicted quality variable to identify subgroups.** Laying aside the modeling sample from this point on, the remainder of the treatment group and the entire control group are assigned a measure of their predicted quality subgroup membership, based on which category (0, 1 or 2) they have the highest probability of belonging to, given their baseline characteristics.

**Step 4. Analyze the impact of quality by comparing the treatment and control groups’ mean outcomes, by subgroup.** Although this kind of analysis can involve a conventional split-sample subgroup analysis, we will follow the HSIS’s existing practice of pooling data and computing subgroups’ impact estimates accordingly.

**Step 5. Convert impacts for predicted quality subgroups to impacts on actual quality subgroups.** This final step converts the impact estimates from Step 4, which represent impacts on predicted subgroups, to represent impacts on actual subgroups, under certain assumptions.

To design the conversion process, we begin with three equations that posit that the impact on each of the three predicted subgroups (no-shows, low quality participants, and high quality participants, respectively) is a linear combination of the impacts on actual subgroups, where the
weights on each element are the proportion of the subgroup that are classified into that group (see Appendix Table B1 for notational definitions).

\[ I_N = s_N N_N + w_N L_N + g_N H_N \]
\[ I_L = s_L N_L + w_L L_L + g_L H_L \]
\[ I_H = s_H N_H + w_H L_H + g_H H_H \]

This set of three equations contains nine unknowns, and so some (six) assumptions are necessary in order to solve the system. In this application, we make the following six assumptions:

1. \( N_N = 0 \) – the impact on predicted no-shows who are actual no-shows is zero
2. \( N_L = 0 \) – the impact on predicted low quality participants who are actual no-shows is zero
3. \( N_H = 0 \) – the impact on predicted high quality participants who are actual no-shows is zero
4. \( L_H = L_L \) – the impacts on low quality participants are the same for children predicted to be high quality participants and children predicted to be low quality participants
5. \( H_H = H_L \) – the impacts on high quality participants are the same for children predicted to be high quality participants and children predicted to be low quality participants
6. \( H_N = L_N = H_L = L_L \) – the impact on high quality participants differs from impact on low quality participants by the same amount whether one looks at high and low quality cases predicted to be no-shows or high and low quality cases predicted to be low quality participants

Ultimately, we rearrange the equations above, imposing our assumptions, to express the terms of interest—impacts on the actual subgroups—as a function of the elements that are known, the impacts on predicted subgroups and the relative proportions of those predicted to be in each group who are actually in each group (conversion factors appear in Appendix Table B2). This conversion process is necessary because some misclassification of children is inevitable—e.g., predicting a child who actually receives low quality Head Start as likely to receive high quality Head Start. The paper provides evidence from the data that support the credibility of the necessary assumptions for translating the internally valid impact estimates for predicted quality subgroups into more policy relevant impact estimates for actual quality subgroups. It also discusses alternative assumptions and the implications of the varied assumptions for the robustness of analytic results.

Usefulness / Applicability of Method:

Demonstration of the usefulness of the proposed methods using hypothetical or real data.

In brief, this method can be extended to other “what works” social policy questions in random assignment studies.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

This research, while providing important information about the role of quality in generating Head Start’s impacts, focuses here on the methodological innovations developed and applied to do so. Therefore, the findings and conclusions for the SREE audience are methodological in nature, considering how to evaluate causal mediators in the context of social experiments.
Appendices
Not included in page count.

Appendix A. References

References are to be in APA version 6 format.


Table B1. Definition of Notation Used in Conversion Equations

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_N$</td>
<td>impact on predicted no-shows</td>
</tr>
<tr>
<td>$I_L$</td>
<td>impact on predicted low quality participants</td>
</tr>
<tr>
<td>$I_H$</td>
<td>impact on predicted high quality participants</td>
</tr>
<tr>
<td>$N_N$</td>
<td>impact on predicted no-shows who are actual no-shows</td>
</tr>
<tr>
<td>$N_L$</td>
<td>impact on predicted low quality participants who are actual no-shows</td>
</tr>
<tr>
<td>$N_H$</td>
<td>impact on predicted high quality participants who are actual no-shows</td>
</tr>
<tr>
<td>$L_N$</td>
<td>impact on predicted no-shows who are actual low quality participants</td>
</tr>
<tr>
<td>$L_L$</td>
<td>impact on predicted low quality participants who are actual low quality participants</td>
</tr>
<tr>
<td>$L_H$</td>
<td>impact on predicted high quality participants who are actual low quality participants</td>
</tr>
<tr>
<td>$H_N$</td>
<td>impact on predicted no-shows who are actual high quality participants</td>
</tr>
<tr>
<td>$H_L$</td>
<td>impact on predicted low quality participants who are actual high quality participants</td>
</tr>
<tr>
<td>$H_H$</td>
<td>impact on predicted high quality participants who are actual high quality participants</td>
</tr>
<tr>
<td>$s_N$</td>
<td>proportion of predicted no-shows who are actually no-shows</td>
</tr>
<tr>
<td>$s_L$</td>
<td>proportion of predicted low quality participants who are actually in the no-show subgroup</td>
</tr>
<tr>
<td>$s_H$</td>
<td>proportion of predicted high quality participants who are actually in the no-show subgroup</td>
</tr>
<tr>
<td>$w_N$</td>
<td>proportion of predicted no-shows who are actually in the low quality subgroup</td>
</tr>
<tr>
<td>$w_L$</td>
<td>proportion of predicted low quality participants who are actually in the low quality subgroup</td>
</tr>
<tr>
<td>$w_H$</td>
<td>proportion of predicted high quality participants who are actually in the low quality subgroup</td>
</tr>
<tr>
<td>$g_N$</td>
<td>proportion of predicted no-shows who are actually in the high quality subgroup</td>
</tr>
<tr>
<td>$g_L$</td>
<td>proportion of predicted low quality participants who are actually in the high quality subgroup</td>
</tr>
<tr>
<td>$g_H$</td>
<td>proportion of predicted high quality participants who are actually in the high quality subgroup</td>
</tr>
<tr>
<td>$1-r$</td>
<td>proportion of low quality participants who are predicted as no-shows</td>
</tr>
<tr>
<td>$1-p$</td>
<td>proportion of high quality participants who are predicted as no-shows</td>
</tr>
</tbody>
</table>

Table B2. Conversion Factors

Impact on Low Quality Participants

\[
L = \left( \frac{1 - r}{w_N + g_N} \right) I_N - \left[ \frac{(1 - r)g_N(w_H + g_H) + rg_H(w_N + g_N)}{(w_Hg_L - w_Lg_H)(w_N + g_N)} \right] I_L
+ \left[ \frac{(1 - r)g_N(w_L + g_L) + rg_L(w_N + g_N)}{(w_Hg_L - w_Lg_H)(w_N + g_N)} \right] I_H
\]

Impact on High Quality Participants

\[
H = \left( \frac{1 - p}{w_N + g_N} \right) I_N - \left[ \frac{(1 - p)w_N(w_H + g_H) + pw_H(w_N + g_N)}{(w_Hg_L - w_Lg_H)(w_N + g_N)} \right] I_L
+ \left[ \frac{(1 - p)w_N(w_L + g_L) + pw_L(w_N + g_N)}{(w_Hg_L - w_Lg_H)(w_N + g_N)} \right] I_H
\]

SREE Spring 2013 Conference Abstract Template