

## **Symposium Justification:**

This symposium unites four papers focused on different aspects of the University of Pennsylvania's 3-year, mixed-methods evaluation of *Zoology One: Kindergarten Research Labs*, an integrated literacy and science curriculum for kindergarten that was developed by American Reading Company. The research is funded by a Goal 3 efficacy grant from the U.S. Department of Education's Institute for Education Sciences, and is conducted in the School District of Philadelphia. The study includes a multi-site cluster-randomized controlled trial with 80 classrooms and roughly 2,000 kindergarten students over 2 year-long cohorts with longitudinal follow-up. It includes cost and implementation studies, and both traditional and non-traditional student outcomes. It is entering its third and final year.

Prior research has demonstrated the promise of instructional approaches that integrate literacy with a content-area focus (Slavin et al., 2014; Cervetti et al., 2012; Goldschmidt & Jung, 2011; Guthrie et al., 1999a; Guthrie et al., 2004; Pearson, Moje, & Greenleaf, 2010; Romance & Vitale, 2001; Shanahan et al., 2010; Wang & Herman, 2005; Wigfield et al., 2008). However, this prior research has focused on older students. The current study examines the impact of this integrated approach on kindergarten students' learning in literacy as well as science, thereby offering contributions to the research base on how young children can learn best in both of these critical content areas.

The *Zoology One* study offers additional contributions, including a mixed-methods examination of hypothesized mediators, such as motivation to read; an integrated and methodologically novel cost study; the development of a new online science assessment for kindergarten that can be used by teachers as well as other research teams; and a focus on non-traditional outcomes such as student learning behaviors. This symposium will present the methods and findings of this research, with an emphasis on the trade-offs and tensions resulting from both the unusually comprehensive scope of the evaluation--spanning traditional and non-traditional student outcomes--and the challenging evaluation context in a large urban school district. The study offers new insights on critical topics to program evaluation generally, including the scope and methods of the cost study; the approach to addressing joiners to the RCT and the particular joiner issues that arise when participants are in kindergarten; and the process of creating and/or adapting some outcome measures to effectively assess a young and historically underserved student population.

**Paper 1 Title:** Measuring academic outcomes in a hard-to-assess student population

**Authors:** Dubois, T., Gray, A., Suwak, K. Sirinidies, P.

**Presenting Author:** Katarina Suwak, University of Pennsylvania

**Background/Context:** The standardized assessments typically used by evaluators to measure academic outcomes are developed using norming samples that represent the US as a whole, making them widely applicable. However, national norming samples do not necessarily represent historically underserved urban students well, limiting the utility of even widely used standardized assessments for studies in challenged urban settings. As an additional challenge, good measures of some academic outcome domains (e.g., science and writing) are difficult to find for kindergarten students at all.

**Purpose/Objective/Research Question:** This paper describes the challenges the evaluation team encountered in seeking appropriate measures of academic outcomes for the young and generally low-achieving participants in the *Zoology One* study. Topics will include our process of first adapting an existing science assessment and then ultimately developing and piloting our own instrument; and the development of enhanced scoring guidelines for a standardized and widely used writing assessment--the Writing subtest of the Kaufman Test of Educational Achievement, Third Edition (KTEA-3)--in order to address its deficiencies for our study's sample.

**Setting:** School District of Philadelphia.

**Population/Participants/Subjects:** The *Zoology One* study includes approximately 2000 students attending kindergarten in the School District of Philadelphia, where 77 percent of students are from historically underserved racial groups and more than 90% qualify for free lunch.

**Intervention/Program/Practice:** *Zoology One* is a year-long curriculum that fully integrates literacy and science learning in kindergarten. It consists of a 90-minute instructional block that adheres to a balanced literacy framework, including morning meeting, readers' and writers' workshops using science-based texts and materials, small-group instruction, and hands-on science exploration activities. In the current study, *Zoology One* is implemented in place of regular literacy instruction in kindergarten. To guide implementation, teachers receive startup professional development and ongoing classroom-based coaching from professional coaches employed by the program's developer, American Reading Company. The program emphasizes independent reading in leveled texts, home reading, ongoing formative assessment, and daily writing and science instruction.

**Research Design:** To construct the science measure, a large pool of items was developed to reflected the Next Generation Science Standards. Items were first administered to teachers and students to ensure feasibility and clarity. More than 50 items were then piloted in non-RCT schools. Piloted items were administered in 16 booklets each with 20 items the combinations determined by a design matrix for balanced incomplete booklets using orthogonal spiraling optimized for maximum spread. Two-parameter IRT analyses of the pilot data informed

selection of 21 highly discriminating items with maximum spread of difficulty for all standards. The resultant measure was administered to the RCT population in spring as an exploratory outcome measure. Psychometric properties are reported for final administration including scale dimensionality and internal consistency.

Adaptation of the Kaufman scoring guidelines began as an effort to clarify scoring requirements, but grew to include the creation of a companion scoring manual. This manual was designed to help the scorers on the *Zoology One* project navigate scoring dilemmas that resulted from the low writing proficiency of the study participants, and that were not adequately addressed by the published guidelines. Scorers received extensive training and the research team used a rigorous and continuous reliability process to ensure that all scoring guidelines were applied consistently.

**Data Collection and Analysis:** Both the science and writing assessments were administered to kindergarten students individually by trained assessors with established reliability. The science assessment was administered on a tablet using the Qualtrics application. Results of both pilot and final administrations were analyzed using item response theory. In contrast, the writing assessment was delivered on paper and required four months of time for seventeen team members to be trained, reach reliability, and score the assessments. Data analysis for the writing assessment is underway.

**Findings/Results:** The paper will describe our newly created science assessment, the Early Life Sciences Assessment (DuBois, Gray, Sewak, Sirinides, 2017) and its development process in depth, and will detail our adaptation of the Kaufman scoring guidelines to better suit our student population. Key findings from the two assessments will be discussed (analyses are underway at the time of this submission). In addition, we will address the benefits and drawbacks of our modification/creation of measures for the evaluation, and implications of this process and our findings for future studies.

**Conclusions:** While it is advisable to use developed and standardized measures wherever possible, in some cases appropriate measures may not be available and alternative instruments may need to be created or modified to fit the population. Our study offers insights and lessons for these situations.

**Paper 2 Title:** Assessing non-traditional outcomes in young children

**Authors:** Fink, R., Gray, A., Dubois, T.

**Presenting Author:** Ryan Fink, University of Pennsylvania

**Background/context:** The evaluation of *Zoology One* includes student-level outcomes in reading, writing, and science. However, in order to comprehensively test the program theory, the evaluation team also sought to examine impacts on less traditional outcome domains. Specifically, we were interested in understanding *Zoology One*'s impacts on young students' motivation to read--a hypothesized mediator for the impacts of the curriculum on academic outcomes--as well as on a cluster of outcomes we call "learning behaviors," which includes independence, confidence, verbal expression, and engagement. Because no appropriate measures of these particular outcomes existed for our study's population, we created measures for use in this research.

**Purpose/Objective/Research Questions:** This paper will detail both the need for and the research team's process of developing measures of motivation to read in kindergarten and a set of learning behaviors identified by the program theory. We will share findings from the administration of both measures and the implications of this work for the study and future research.

**Setting:** School District of Philadelphia.

**Population/Participants/Subjects:** The *Zoology One* study includes approximately 2000 students attending kindergarten in the School District of Philadelphia, where 77 percent of students are from historically underserved racial groups and more than 90% qualify for free lunch. This paper will discuss the administration of the Kindergarten Reading Motivation Scale (KRMS, Castillo & Gray, 2017) administered to 951 students across 12 schools in March 2017, and the administration of a teacher-completed assessment of learning behaviors for former study participants when they reach first and second grades.

**Research Design:** The self-report, individually administered KRMS was developed and piloted via a rigorous instrument-development process in response to the dearth of appropriate reading motivation measures for young children. It focuses on two domains: 1) Perceived self-efficacy, and 2) Desire to read. The classroom learning behavior survey was developed in response to program theory that suggested the intervention would have impacts in non-academic domains. Potential areas to examine were then vetted with *Zoology One* curriculum experts and coaches. Teachers were asked to respond to items about students' classroom learning behaviors, including independence, confidence, verbal expression, and engagement.

**Data Collection and Analysis:** The reading motivation data was collected by trained assessors and scored by the research team. Correlations with reading achievement measures were calculated. The online classroom learning behavior survey was sent first- and second-grade teachers of students who participated in the study during kindergarten.

**Findings/Results:** Preliminary results from the KRMS reveal statistically significant treatment impacts on students' self-reported motivation. The correlation with reading achievement was not statistically significant. As an additional area of inquiry, we conducted an analysis of kindergarten students' likelihood of identifying gender-stereotypical (Bussey & Bandera, 1992, 1999) "favorite books." This analysis revealed that girls in ZOO classrooms were *more likely* than girls in control classrooms to identify "favorite books" that were non-gender-stereotypical. (e.g., they were more likely to identify favorite books about topics such as animals, sports, or vehicles as opposed to books about princesses or Barbie). The difference was statistically significant.

The *classroom learning behavior* survey data were still being analyzed at the time of this submission. Results will be shared during this presentation.

**Conclusion:** Instruments designed to assess non-traditional outcomes for kindergartners are not readily available, but assessing these outcomes is critical to developing a full understanding of the impacts of a curricular intervention. This paper offers insights on the process of identifying areas where non-traditional outcomes may be important, and how our team developed instruments for doing so. It includes both the contributions of these efforts to our study and potential limitations.

**Paper 3 Title:** Inclusion of costs and the tradeoff between rigor and purchase price

**Authors:** Bowden, A.B, Rodriguez, V.

**Presenting Author:** A. Brooks Bowden, North Carolina State University

**Background/Context:** Randomized field trials are conducted to provide information about a program's effect on an outcome of interest. The primary focus of the research within an evaluation is estimating impacts. However, within the evaluation, it is also critical to examine the processes and resources (also called "ingredients") utilized during implementation in order to understand the contrast between the treatment and control groups in the production of the outcome (Levin, McEwan, Belfield, Bowden, & Shand, 2018; Century & Cassata, 2016; Hulleman & Cordray, 2009). In the same way that an RCT provides internally valid causal estimates of the impact of the program, a cost analysis following the ingredients method provides an exact mapping of the resources used in producing the impacts observed.

**Purpose/Objective/Research Question:** This paper focuses on the tradeoff between capturing all costs, such as volunteer time, reallocated staff time or school resources, and parental time, and reflecting only the budgetary commitment required to purchase the program of interest. This tradeoff or tension is illustrated through the ongoing evaluation of an early reading program, *Zoology One*.

**Setting:** School District of Philadelphia.

**Population/Participants/Subjects:** The *Zoology One* study includes approximately 2000 students attending kindergarten in the School District of Philadelphia, where 77 percent of students are from historically underserved racial groups and more than 90% qualify for free lunch.

**Intervention/Program/Practice:** *Zoology One* is a year-long curriculum that fully integrates literacy and science learning in kindergarten. It consists of a 90-minute instructional block that adheres to a balanced literacy framework, including morning meeting, readers' and writers' workshops using science-based texts and materials, small-group instruction, and hands-on science exploration activities. In the current study, *Zoology One* is implemented in place of regular literacy instruction in kindergarten. To guide implementation, teachers receive startup professional development and ongoing classroom-based coaching from professional coaches employed by the program's developer, American Reading Company. The program emphasizes independent reading in leveled texts, home reading, ongoing formative assessment, and daily writing and science instruction.

The program is guided by the theory that children learn better when they are motivated and engaged, when they read frequently and with different individuals, when materials are targeted at their reading level, and when reading occurs in the classroom and at home. This theory hinges on the availability of materials and books to read, highly trained and prepared teachers, and a

curriculum that comprehensively addresses literacy development. An additional component to this process is the incorporation of authentic assessment and data driven instruction.

The program was delivered in two cohorts where approximately half the sample was served in year one and the other half was served in year two. Implementation for the evaluation was completed in June 2018.

**Research Design:** Costs are estimated following the ingredients method (Levin, McEwan, Belfield, Bowden, & Shand, 2018). It is important to note that because data collection on resource use is done simultaneously with data collection on implementation, in this study we are able to estimate costs for each classroom in each school. Because most cost analyses are generally done retrospectively and because resource information is extremely challenging to recover, most cost analyses on education have relatively small sample sizes and are subject to the usual objections that come with it. Therefore, this study provides a novel large dataset on costs on education that is seldom seen in the literature.

**Data Collection and Analysis:** The cost component of the evaluation rely upon program management information system data, school district data, classroom observations, interviews, teacher surveys, and teacher time logs.

**Findings/Results:** While challenging, it is critical that the cost analysis reflect all resources used to deliver the program and produce the outcome. This allows the work to contribute to the understanding of the program's impact and to provide the necessary information to replicate the program and its impacts successfully. In addition to describing and estimating total costs, it is critical to also address the question of how the resources are financed across multiple agents. This allows the analysis to clarify which resources were contributed by outside agents, such as parents. However, if the parental contribution was excluded the results may not be replicated successfully given that this resource was non-trivial and part of the program's theory of change.

**Conclusions:** This paper recommends that cost analyses apply the ingredients method to document and describe costs from a total social perspective and to supplement this analysis with an examination of how those costs were financed.

**Title:** Kindergarten in SDP: A special case of joiners

**Authors:** Sirinidies, P, Gray, A.

**Presenting Author:** Philip Sirinides, University of Pennsylvania

**Background/Context:** The impact of joiners--sample members who join the study after random assignment--on a rigorous study is much discussed in the program evaluation literature and has been address by the What Works Clearinghouse group design standards over multiple revisions. Joiners threaten internal validity because of the risk that these individuals may have intentionally selected or been placed into one condition or the other after assignment, thereby invalidating the random process. The joiners issue raises special questions when the study participants are kindergarten students beginning school in a challenged urban district. In Philadelphia, for instance, typically fewer than 50% of the students who ultimately attend kindergarten in SDP schools are registered by the first day of school. Of the nearly 13,000 children who were enrolled in SDP kindergarten for 2016-17, only 44% were registered when staff left for summer recess. Twenty-three percent of the kindergarten class was still unregistered a week into the school year, and roughly 15% of the class had not yet registered by the end of the second week of school (SDP, 2016). As a study of kindergarten students in this context, the *Zoology One* study is therefore particularly challenged to address the issue of joiners, and is well positioned to offer insights on this issue.

**Research Design:** The evaluation design for the IES-funded efficacy trial of Zoology One includes a rigorous experimental research design that supports strong causal inferences about program impacts. Cluster randomization was selected because Zoology One is a whole-class intervention. Randomization was conducted at the classroom level with entire classrooms—including teacher and all students—in September, after students and teachers have been assigned to classrooms. To determine sample status, class rosters were collected before random assignment and compared to students in attendance at the time of fall assessment (baseline). This paper focuses on the baseline equivalence of students in the stable sample and early joiners; and explores the moderating effects of being an early joiner on program impacts.

**Data Collection and Analysis:** Differences at baseline between early joiners and students in the stable sample on WRMT-III's Reading Readiness Cluster are estimated using a three-level hierarchical linear model, with students nested within classrooms and classrooms nested within schools. The treatment-control differences are estimated separately for the full and stable samples using the same three-level HLM with pretest scores as a covariate and a treatment status indicator for the impact of Zoology One.

**Findings/Results:** Description of main findings with specific details. Any assessed student who was listed on an original class roster was considered part of the stable sample. Two schools did not provide rosters prior to random assignment. In the first year, forty eight percent of assessed students were in the stable sample, 23% were in schools that did not provide rosters, 27% were individual early joiners, and 2% changed classrooms after assignment in ways that crossed experimental condition. While baseline equivalence was established for treatment and control groups, students who were in the stable sample score .29 SD higher on WRMT-III's Reading

Readiness Cluster at baseline. Analyses of the stable sample revealed overall significant effects of Zoology One, including a treatment effect of .28 standard deviations for the WRMT passage comprehension subscale. However, no significant differences were found when analyzing data from the full sample.

**Conclusions:** In a cluster randomized control trial, participants who enter the study after group assignment will weaken causal inferences about program impacts. This study details the research team's process for exploring the prevalence and inclusion of joiners in the context of a kindergarten RCT study, and the implications of these decisions for the study's internal validity.