A Guide to Conducting School Experiments: Expert Insights and Best Practices for Effective Implementation

Kerstin Grosch*[1], Simone Haeckl†[2], Holger A. Rau‡[3,4], and Paul Preuß§[4]

[1]WU Vienna University of Economics and Business
[2]University of Stavanger
[3]University of Duisburg-Essen
[4]University of Göttingen

January 30, 2024

Abstract

This guide provides a comprehensive overview of the distinct characteristics of school experiments conducted with children in preschools and schools. We investigate and describe the essential considerations involved in designing and implementing such experiments, drawing insights from a survey with senior researchers. Moreover, the guide summarizes nine key lessons learned from the experiences of these researchers. The paper also presents the opinions of inexperienced researchers in school experiments (juniors) on crucial aspects of successful school experiments, which differ from the opinions of the experienced senior researchers. As a result, this guide serves as a valuable resource for junior researchers embarking on their first school experiments. Promoting the adoption of best practices endorsed by senior researchers (e.g., ethical considerations, age-appropriate experimental procedures), the guide strengthens the success, validity, and reliability of future school experiments.

JEL Classification numbers: B41, C93, I20

Keywords: school experiments, guide, internal validity, survey

*Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria; *email: kerstin.grosch@wu.ac.at*

†University of Stavanger, Business School, Kjell Arholms gate 37, 4021 Stavanger, Norway; *email: simone.haeckl-schermer@uis.no*

‡University of Duisburg-Essen, Universitätsstraße 2, 45141 Essen, Germany; *email: holger.rau@uni-goettingen.de*

§University of Göttingen, Wilhelmsplatz 1, 37073 Göttingen, Germany; *email: paullukas.preuss@stud.uni-goettingen.de*

We acknowledge the financial support provided by the Joachim Herz Foundation and the Austrian Science Fund (FWF) under grant number [T 1263-G]. We are grateful for the valuable and comprehensive feedback provided by Maria Polipciuc and Mari Rege on the initial version of this paper. We also express our appreciation to Gary Charness, Catherine Eckel, Henning Hermes, Fabian Kosse for their helpful comments. We would like to extend our gratitude to Ankush Asri, Pablo Brañas-Garza, Catherine Eckel, Seda Ertac, Daniel Martínez Felip, Amalia Di Girolamo, Mofei Jia, Luisa Lorè, Maria Polipciuc, Benjamin Prissé, Eva Ranehill, and other anonymous participants who contributed to the surveys.

1 Introduction

Field experiments have a relatively long history in economics, reaching back to at least the 1960s (Dufwenberg et al., 2008). More recently, school experiments, i.e., experiments conducted with children in preschools and schools, have gained in importance in economics. Similar to other experiments with children, school experiments allow for analyzing the evolution of preferences, attitudes, and behavior in early life (e.g., Angerer et al., 2015; Bindra et al., 2020; Brocas and Carrillo, 2020; Harbaugh and Krause, 2000; Schunk and Zipperle, 2023; Sutter and Glätzle-Rützler, 2015; Sutter et al., 2019). Moreover, school experiments can test interventions aiming at reducing achievement and opportunity gaps with a control and treatment group. Frequently, these experiments strive to enhance educational outcomes, non-cognitive abilities, or better overall choices among children (e.g., Alan et al., 2021; Cappelen et al., 2020; Charness et al., 2023, 2019; Grosch et al., 2022; Kosse et al., 2020; Ozturk et al., 2020). School experiments encompass a range of approaches. In this guide, we focus on lab-in-the-field experiments which investigate children's preferences and behavior in a controlled environment and intervention studies which experimentally evaluate programs. The successful implementation of school experiments requires not only the researchers' professional competence but also awareness of the unique characteristics of the school context. This is crucial to achieving high implementation fidelity, i.e., a high degree to which an intervention or experiment is delivered as intended (e.g., Carroll et al., 2007), compliance with ethical standards, and building lasting relationships with stakeholders (e.g., headmasters, school authorities, teachers).

In this paper, we identify and describe the distinctive aspects of school experiments compared to other field experiments. We surveyed experts in the field to identify key aspects for designing and conducting school experiments. Our objective is to raise awareness about these aspects and offer a set of guidelines for best practices in conducting school experiments. In an online survey, we asked experts to state (i) the most important aspects to consider when conducting school experiments, (ii) the lessons they learned in past school experiments, and (iii) how to build lasting relationships with stakeholders. We grouped the responses (i) into seven categories/aspects, enriched with literature and descriptions. We summarize the responses to questions (ii) and (iii) in nine practical lessons for navigating successful school experiments and stakeholder relationships. We conducted a second survey with junior researchers. Specifically, we examined whether juniors accurately anticipate the views of experts, which is crucial for meeting academic requirements, increasing publication success, and successfully implementing school experiments. To accomplish this, we presented the categories/aspects identified by the experts to the juniors, randomized them into two groups, and asked them to either predict the experts' ranking of these aspects (i.e., their second-order beliefs) or state their own preferred ranking. Our findings reveal a divergence between second-order beliefs and the experts' opinions, highlighting that juniors do not accurately anticipate experts' opinions. In addition, we find a notable difference in juniors' and seniors' own preferred rankings, with seniors prioritizing experimental procedures and ethical considerations, while juniors value an age-appropriate design and good relationships with partners more. Therefore, juniors may have different priorities than seniors when planning their first school experiments, potentially warranting the transmission of knowledge from experts to novices in the field. This expert-driven may fulfill this role and serve as a valuable resource for researchers in the field of school experiments.

Our guide makes several contributions to the literature. First, it identifies and describes the distinct aspects of school experiments, setting the base for best practices for effective implementation, high implementation fidelity, compliance with ethical standards, and building lasting relationships with stakeholders. Second, it highlights the important aspects of school experiments, aiding novices in school experiments to navigate through the process of conducting school experiments. Finally, the guide raises awareness of the necessary resources, e.g., budget, time, and information, required for school experimental research which may be underestimated without prior experience in the field.

This guide has limitations in its scope. It does not intend to cover every aspect distinguishing school experiments from field experiments. Instead, we focus on the most crucial aspects as identified by experts. For more technical aspects of field-experimental study designs, we mention relevant literature but do not dive into them. We encourage readers to explore those further on their own. Furthermore, this guide does not assist researchers in selecting the appropriate method – whether school experiments are suitable for their research questions or not. When discussing the distinct aspects and practical lessons learned, we refer to the two types of school experiments, lab-in-the-field experiments or impact evaluations, as needed. We do not explicitly discuss combinations of the two types or additional extensions like parent surveys to maintain the guide's comprehensiveness.

The outline of this paper is as follows. Chapter 2 is the core of the paper, where we introduce the senior survey and describe each identified aspect of school experiments from this survey in more detail. Chapter 3 delves into the junior survey, where we explore the perspectives of inexperienced researchers on the distinct aspects of school experiments introduced in the preceding chapter. Chapter 4 synthesizes nine practical lessons derived from the senior survey, providing a roadmap for successfully navigating the intricacies of school experiments. Finally, in the concluding chapter, we summarize and reflect on our findings.

2 Important aspects of school experiments

In this section, we first outline the senior survey and our process for identifying the key aspects of school experiments. Then, we delve into each of these aspects individually, describing and discussing them in separate subsections.

2.1 Senior Survey

To gain insights into the crucial considerations for designing and conducting school experiments, we conducted a survey among senior researchers in the field. The survey consisted of two parts. In the first part, experts were asked to identify the three most significant aspects to consider when implementing experiments in schools. The responses, collected through an open text box, were analyzed and categorized, as elaborated in Subsection 2.2. The second part of the survey focused on obtaining practical advice to establish enduring relationships with stakeholders and their lessons learned from previous school experiments. We present these insights as nine practical lessons in Section 4.

2.1.1 Recruitment and sample description

The senior researchers were chosen if they had published at least one school experiment in a peer-reviewed journal. We only considered studies that involved data collection in schools, preschools, or kindergartens. In addition, we specified that at least one of these studies had to be conducted in western educated industrialized rich democratic countries (WEIRD countries) to increase comparability between the target countries of the studies.[1] We contacted 227 senior researchers that met these criteria and were either from the field of economics or psychology (of those who responded, only three were not economists).[2] We sent them personalized emails and invited them to participate in a 10 to 15-minute online survey on Qualtrics in June 2022. 25 senior researchers completed the survey.

Most of the senior researchers in our sample hold the academic title of "Full Professor" (60%), followed by "Assistant Professor" (8%), and "Post-Doc" (12%).[3] The majority of the participants (88%) indicated "Economics" as their main research discipline. Most of the senior researchers work at universities (84%). Most of the senior researchers have extensive experience in conducting experiments in schools, i.e., 52% of the respondents conducted five or more, and 25% conducted at least 10 experiments in schools. Moreover, 70% of the respondents have more than ten peer-reviewed publications, with eight having over 40 publications.

2.1.2 Results

In total, 25 senior researchers finished the survey and provided 60 answers that we used for the analysis.[4] We applied an inductive approach, grouping the answers into categories in several discussion rounds with all authors. We made sure that the categories covered most of the answers received. We identified seven different categories (frequency of mentions in parentheses) "age-appropriate experimental procedures" (16), "ethical considerations" (13), "age-appropriate experimental design" (10), "support from partners" (10), "age-appropriate incentives" (5), "timing within school year and day" (4), and "statistical power" (2). In the following section, we will describe each category in greater detail.

[1] Compared to WEIRD countries, non-WEIRD countries may differ concerning technical and logistical infrastructure, availability of research partners, and participants. We decided to focus on WEIRD countries since the majority of school experiments are conducted in WEIRD countries and the authors of this paper are experienced in conducting school experiments in WEIRD countries and, thus, can contribute with their expertise to this guide. We hope that the guide is still useful for all researchers.

[2] Two seniors state that their main field is Psychology and one that it is Sociology. The results presented below do not change when we exclude these three seniors.

[3] 20% of the respondents answered the question about their academic rank with "other" or "prefer not to say."

[4] Some respondents listed less than three different aspects, reducing the number of answers to 60 instead of 75.

2.2 Aspects

In this section, we describe the unique aspects of school experiments in more detail.[5] In the folllowing, we refer to individuals under the age of 18 as "children." Each subsection refers to a category identified by the senior survey (see subsection 2.1.2). We organize the various aspects starting with the description of those integral to the initial planning stages of a school experiment and progressing to aspects more closely associated with practical implementation. However, it is important to note that the different aspects are interconnected and also not strictly distinguishable from each other. Moreover, planning and conducting a school experiment involves navigating between various aspects, as changes in one area may necessitate adjustments in another, creating a cyclical process.

2.2.1 Experimental design

Considering the developmental stage of children is critical to ensure their comprehension of experimental tasks. This principle applies not only to lab-in-the-field experiments, but also to outcome measures used in program evaluations. Depending on the specific outcomes of interest, researchers may have access to validated and age-appropriate measures. For instance, standardized tests are viable options for assessing academic skills and executive functions. Additionally, survey constructs can effectively evaluate essential skills and non-academic outcomes, having undergone validation processes tailored to specific age groups.[6] However, economists generally prefer measuring actual choices and behaviors over stated preferences or behaviors obtained through surveys. The majority of experiments measuring revealed preferences have primarily been designed for laboratory settings, utilizing samples comprised of university students. Therefore, adaptations are necessary to apply these experiments in field settings (Eckel and Candelo Londono, 2021). This need for adaptation becomes particularly pertinent when conducting experiments involving children, given their potential lack of fully developed essential competencies. These competencies include academic skills (e.g., literacy and numeracy) and executive functions (e.g., working memory, inhibitory control, and attention shifting, which refers to the ability to switch tasks in response to changing circumstances) (List et al., 2023). To ensure effective comprehension among children, several adjustments must be made in the experimental set-up. For instance, the language employed in instructions should be tailored to the appropriate age group, and the mode of instruction delivery should be adapted accordingly, such as reading the instructions aloud rather than relying on the children to read independently. This approach minimizes the potential impact of varying literacy skills on the results, thereby reducing heterogeneity. Additionally, incorporating visualizations can facilitate task understanding, particularly when explaining experimental tasks, random processes (e.g., Guerini et al., 2020; Harbaugh et al., 2002; Piovesan and Willadsen, 2021; Tobol and Yaniv, 2019), probability distri-

[5] For a comprehensive overview of the benefits and the conduct of field experiments in general, we recommend the paper by List (2011).

[6] See, e.g., the SPECTRUM database for an overview of psychometric measures https://educationendowmentfoundation.org.uk/spectrum-essential-skills-and-non-academic-outcomes/spectrum-database

butions, allocation decisions (e.g., Alan et al., 2020; Andreoni et al., 2020; Bauer et al., 2014; Ben-Ner et al., 2017; Choshen-Hillel et al., 2020), or cooperation decisions (e.g., Sutter and Untertrifaller, 2020). For instance, the use of an urn can effectively illustrate probability distributions when measuring risk preferences (Alan et al., 2017). Furthermore, payoffs in games involving strategic interactions, like the public-good game, can be illustrated using coins. These coins can be allocated to a ”private basket” to represent individual gains or multiplied in front of the children before being placed into a ”public basket” to symbolize shared benefits (Hermes et al., 2020).

To enhance children’s engagement, experimental games must be interesting, simple, and tailored to the specific context. Numerous established experimental methodologies have been successfully adapted for studying children, as summarized in previous works (see, e.g., List et al., 2023; Sutter et al., 2019, for summaries). A notable feature of these adaptations is their emphasis on minimizing the requirement for abstract thinking, which is a competency that develops with age. One effective strategy to reduce abstractness is to utilize tangible tokens instead of abstract units. For instance, when measuring time preferences, employing physical tokens that children can place into an envelope if they wish to save them for later has proven helpful (Sutter et al., 2015). Similarly, in the context of measuring social preferences, the use of two bowls or envelopes allows children to distribute tokens between themselves and another individual (Gummerum et al., 2010). Additionally, an often-implemented adaptation for eliciting competitive preferences involves engaging children in actual competitions rather than relying on imaginary scenarios, as suggested by (Gneezy and Rustichini, 2004).

List et al. (2023) provides a comprehensive overview of the adaptations made to experimental designs to accommodate the specific needs and competencies of children, such as literacy and numeracy skills. The study examines common practices in school experiments for various age groups, encompassing elements like comprehension checks, session durations, and instructional delivery methods (e.g., electronic, oral, or physical). Moreover, Alfonso et al. (2023) offers a thorough analysis of experimental designs and procedures tailored specifically for adolescents.

2.2.2 Incentives for participants in school experiments

In order to acknowledge participants’ time and effort in experiments, economists commonly provide payment as compensation. Furthermore, economists argue that incentives enhance the reliability and realism of (lab-)experimental outcomes. By incorporating consequences analogous to real-life economic decisions, incentives aid in simulating these effects within the lab setting. Incentives also serve to mitigate demand effects and social desirability bias, thereby improving the quality of data collected. The effectiveness of incentives is contingent upon the nature of the task at hand. For tasks involving individual decision-making, incentives can enhance cognitive and motivational focus, reducing data noise and enhancing internal validity. However, not all tasks necessitate incentives to yield meaningful results (see, e.g., Alfonso et al., 2023; Falk et al., 2022; Voslinsky and Azar, 2021). When assessing the feasibility of implementing incentives, it is crucial to consider factors such as the influence of demand effects and the level of motivation and focus displayed by participants.

When conducting experimental research involving children, alternative forms of incentives beyond monetary rewards are often necessary (Sutter et al., 2019). For instance, young children may receive tokens instead of money, which can later be exchanged for small gifts like stickers or toys at an experimental shop post-experiment. To ensure the salience of incentives, these gifts are typically displayed before the experiment commences (Harbaugh and Krause, 2000). For a comprehensive overview of the most commonly employed incentives in experiments with children across different age groups, we refer to List et al. (2023). When monetary incentives are utilized, their magnitude usually aligns with the average amount of pocket money received by children in the respective age group within the sample (e.g., Fehr et al., 2013). Using the same amount of pocket money in different studies helps ensure comparability. On the other hand, if we don't consider the incentives of a reference study, it can alter the results and make it challenging, if not impossible, to identify any differences from the reference paper.[7] Incentives should not only motivate children but also appeal to their teachers and parents/legal guardians, especially if the children are below a certain age. Hence, it is beneficial to consider the specific characteristics of the subject pool and the preferences of their guardians. For example, certain incentives like candy may be deemed inappropriate for certain subject pools. A useful starting point for selecting incentives is to consult with teachers or headmasters, as they possess knowledge regarding the preferences and appropriateness of gifts for children in specific age groups. Furthermore, incentives should be equitable. In cases where children in a class exhibit significantly disparate skills for the incentivized tasks (e.g., in integrated schools), exceptions may need to be made or incentives for that particular task should be removed altogether due to ethical considerations.

Moreover, incentives should be meticulously calibrated. If the incentives are set too low, they may become insignificant and potentially result in decreased effort levels due to motivational crowding out (e.g., Gneezy et al., 2011; Voslinsky and Azar, 2021). Conversely, if the incentives are set too high, unintended consequences can arise, including performance deterioration under pressure, frustration, and disruptive discussions within the classroom.

Many educators hold the belief that providing monetary rewards to children for academic achievements or effort in learning is morally objectionable. However, the impact of incentives on educational outcomes has been examined in experimental studies, with economists investigating incentives for grades and other educational factors as summarized by Gneezy et al. (2011). In this context, incentives are viewed as the intervention being tested through experimental methods rather than incentives serving as a means to an end as discussed in this subsection. Children are ideally motivated intrinsically to exert effort in learning, and the use of extrinsic rewards may pose a challenge as they have the potential to crowd out long-term motivation (Gneezy et al., 2011). Therefore, when conducting experimental impact evaluations of programs integrated into the school curriculum or implemented during school hours, the provision of extrinsic incentives may not be crucial and could even introduce complications. In fact, it is advisable to refrain from employing incentives to facilitate the generalization of experimental results, as providing incentives may not be feasible on a large scale (Al-Ubaydli et al., 2021). However, for

[7] We thank Fabian Kosse for this remark.

voluntary interventions, non-monetary incentives or behavioral tools such as lotteries rewarding classes with the highest participation rates or regular reminders can be utilized to increase participation rates and reduce drop-outs.

To incentivize schools to participate in an experiment, non-monetary incentives can be employed. These may include providing schools with a concise summary of the research findings and facilitating discussions on the potential implications. In the treatment group, schools may benefit from the intervention, as it aims to enhance an important outcome. However, schools in the control group may miss out on these benefits during the implementation phase of the experiment. By offering the opportunity for schools in the control group to implement the intervention after the official implementation phase, it can serve as a motivational tool to encourage their participation and potentially reduce drop-out rates.

To ensure effective randomization, it is essential to invite all schools to the study under identical conditions. Once the schools have agreed to participate, the randomization of treatments can commence, and each school will be notified of their respective non-monetary incentives based on their assigned group. The overarching objective should be to strike a balance between benefits for social science researchers and to provide value to the schools and children who had a positive experience during and after the experiment, for instance, through debriefing.

2.2.3 Statistical power

In more technical terms, statistical power refers to the probability of rejecting the null hypothesis under the assumption of a specific effect size. Power calculations play a critical role for two key reasons. Firstly, conducting an ex-ante power calculation allows for the assessment of whether the experimental design is suitable for investigating the intended treatment. For instance, if a researcher plans to examine an intervention with an expected small effect size and only has two classes with 25 students available for the experiment, proceeding may not be justified. The likelihood of detecting a small valid treatment effect in a relatively small sample is low, even if individual-level randomization is feasible. Secondly, studies with low statistical power can yield misleading results as the estimated effect size may be biased in both magnitude and direction (Gelman and Carlin, 2014). It is important to emphasize that sufficient statistical power is crucial not only for detecting significant effects but also for accurately identifying null effects. Insufficient power can lead to the oversight of effects that actually exist in the data but fail to reach statistical significance. In other words, low power increases the risk of missing meaningful effects. Hence, researchers should carefully consider the statistical power of their study during the experimental design phase. Through power calculations, they can determine the necessary sample size or make other adjustments to ensure sufficient power for their study. This practice minimizes the risk of drawing inaccurate conclusions and enhances the reliability and validity of the research findings.

The power of an experiment depends on various factors, including the true effect size, the chosen statistical thresholds (e.g., 95% confidence intervals), the number of observations collected, the precision of effect measurement, and the design choices made. As these factors are applicable to all experiments and statistical tests, providing an exhaustive introduction to power calculations is beyond the scope of this discussion (see e.g., Bloom, 1995; Cohen, 2013; Duflo et al., 2007; Glennerster and Takavarasha, 2013). Instead, we will focus on discussing the key aspects specific to design, such as randomization design, that researchers should consider in school experiments and how these factors influence power calculations (See for example Schochet, 2008, for a more detailed and technical discussion of power in educational experiments.).

When considering power calculations, the level of randomization plays a crucial role. The most effective design is individual randomization, where participants are randomly assigned to either the treatment or control group. However, in school experiments, individual randomization is often impractical due to political, organizational, or logistical constraints (Raudenbush et al., 2007; Schochet, 2008). For instance, in a lab-in-the-field experiment, reading instructions aloud might be necessary to improve comprehension, which makes it challenging to implement multiple treatments within a single class. Moreover, interventions pose even greater difficulties for individual randomization. The clustering of children in classes makes it difficult to selectively provide treatment to some children without affecting the non-treated children in the same class. Additionally, well-connected teachers may exchange information about the experiment, posing a threat to internal validity even with class-level randomization within schools. There are additional factors that complicate class-level randomization. Firstly, there may be resistance from teachers who do not receive the intervention, leading to changes in their behavior that could introduce bias into the estimates. In extreme cases, this resistance could even result in the discontinuation of the intervention altogether. Secondly, if teachers communicate with each other, spillover effects may occur between the treatment and control groups. For example, consider an intervention that provides treated teachers with new and exciting teaching materials. It is challenging to prevent teachers from sharing this material with their colleagues, potentially blurring the boundaries between the treatment and control groups. As a result, researchers often opt to randomize at a higher level, such as the class or school level.

However, randomizing at a higher level (cluster) can lead to a reduction in statistical power due to the similarity between individuals within a cluster, which restricts the amount of new information provided by each additional observation within the cluster. An essential factor to consider when using a clustered design is the intracluster correlation coefficient (ICC), which measures the extent to which observations within a cluster are similar. The ICC helps estimate the decrease in power resulting from clustering by indicating the proportion of total variation explained by within-cluster variation. A higher ICC implies that the addition of an extra observation within a cluster provides less benefit. In general, the number of clusters becomes more crucial than the number of observations within each cluster (Raudenbush et al., 2007; Schochet, 2008). Consequently, a well-powered study necessitates a large sample size with numerous schools, which often poses a challenge due to researchers' limited budget.[8] Furthermore, persuading schools to participate in the study can be difficult, resulting in limitations on the available sample size.

[8]Moreover, the cost of adding additional control and treatment clusters (observations) may vary, further complicating the choice of sample size. McConnell and Vera-Hernández (2022) discuss strategies to maximize power in clustered designs with varying costs.

Ultimately, the decision regarding the level of randomization depends heavily on the specific context, including the nature of the intervention and the interactions between treated and non-treated participants, including indirect participants such as teachers, as well as feasibility constraints in terms of sample size.

What can researchers do to improve statistical power? The first step is to determine the lowest feasible level of randomization within the research setting. For example, if randomizing between classes is not feasible due to potential spillover effects, it may still be possible to randomize within grade levels. This approach recognizes that teachers in different grade levels have less interaction, allowing for clustering within grade levels instead of within schools. Moreover, the method of treatment implementation plays a role in determining the level of randomization. While a teacher-implemented intervention often necessitates randomization at the class level, an online intervention may enable individual randomization even when implemented by the teacher.[9]

The second step involves considering baseline covariates to enhance the precision of the estimates, particularly in small samples and clustered designs (Raudenbush et al., 2007). Baseline measures can be used in different ways: they can be utilized to create blocks (or pairs in the extreme case) for stratified randomization, or they can be included as control variables in the analysis. Both approaches improve the precision of the estimates, but stratified randomization has the advantage of guaranteeing balance in important observable variables if the blocking variables are carefully chosen. Simulation studies by Bruhn and McKenzie (2009) demonstrate that stratified randomization is preferable to individual randomization when the sample size is smaller than 300 independent observations. Another approach is to re-randomize until a predetermined level of balance is achieved (Bruhn and McKenzie, 2009). The econometric analysis of treatment effects is influenced by the design choice of randomization methods. A discussion of these methods and their implications can be found in Athey and Imbens (2017).

2.2.4 Ethical considerations

The conduct of scientific research relies on fundamental ethical principles. Experimental economists recognize ethics and morality as captivating research fields that exert influence on human behavior and, consequently, economic outcomes (e.g., Hausman and McPherson, 1993). In their studies, economists adhere to ethical principles, actively opposing deception and ensuring fair compensation for participants' time (see e.g., Charness et al., 2022; Rubinstein, 2001). Furthermore, experimentalists engage in discussions about the ethics of behavioral interventions in public policy, including topics such as paternalism in regulation, welfare effects, and the responsibilities of ethically-driven governments (e.g., Camerer et al., 2003; Chater and Loewenstein, 2023; Sunstein, 2016; Thaler and Sunstein, 2003). Ethical approval, as a prerequisite for conducting field studies like school experiments, is less popular among experimental economists with the potential argument that it hampers scientific progress. However, this perspective stems from a misconception, as research ethics is not intended to impede research but rather to safeguard

[9] An example of this approach can be found in Bettinger et al. (2018), where a clever design utilizing a control group and treatment implementation on computers enables randomization at the individual level.

participants from physical and psychological harm, uphold their human dignity and rights, enhance research quality, and foster trust in research (Resnik, 2018).

Research ethics recognizes children as a ”vulnerable group” due to their status, legal rights, and societal roles, making them more susceptible to harm, exploitation, and potential threats compared to other groups (Ten Have, 2015). As such, special attention must be given to children when they participate in studies. Researchers bear an ethical responsibility to protect children's rights, autonomy, integrity, and dignity. This responsibility is particularly crucial when organizing the consent procedure. Informed consent plays a vital role in ensuring that participants voluntarily engage in research studies. Researchers are obligated to provide participants with comprehensive information about the research content, potential risks and benefits associated with participation, the option to refuse or withdraw from the study, and privacy and data protection considerations prior to obtaining consent. Similarly, children's legal guardians must be informed about these issues and provide their consent before the study can proceed. It is important to note that the specific regulations and ethical standards, as well as the characteristics of the subject pool, including age, may vary across regions, countries, and institutions. Researchers should consult with an ethical committee at their local institution to ensure compliance with the guidelines in the country where their research institute or university is based. In cases where the study is conducted in another country, obtaining ethical approval from both countries is considered best practice. The website of the European Union provides an overview of rules and practices for research with children across different countries, which can serve as a valuable resource.[10] It is essential to present the informed consent in a manner that is accessible and suitable for the target group, taking into account the literacy and comprehension levels of children. The European Commission offers guidance on ethical considerations in social science research, including recommendations on handling informed consents.[11] There are limited exceptions where conducting an experiment without obtaining informed consent may be justified. These exceptions typically occur in natural experiments, as documented in the literature (see e.g., List, 2008).

In ethical considerations for school trials, it is crucial to identify the stakeholders involved, including individuals and institutions with an interest in the study. It is not uncommon for researchers to overlook the importance of debriefing specific stakeholders or providing them with adequate information, such as acknowledging and informing the research staff involved in data collection (Kaplan et al., 2020). The lead researcher of the school trial holds the responsibility to carefully consider the ethical conduct of the study and allocate sufficient time to take appropriate actions.

In the realm of human subjects research, the responsibility for oversight lies with an Institutional Review Board (IRB), also known as Research Ethics Boards (REBs) or Research Ethics Committees (RECs). These boards are typically housed within research institutions and play a critical role in ensuring that studies adhere to established ethical standards. IRBs assist researchers in evaluating the ethics of their research activities and considering potential consequences. They serve as a means to identify ethical

[10] See `https://fra.europa.eu/en/publication/2019/child-participation-research`.

[11] See `https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-in-social-science-and-humanities_he_en.pdf`.

risks for participants and society at large. Researchers can refer to IRB statements to ensure that they have addressed all relevant ethical aspects of their study. The IRB process functions as a valuable tool in recognizing important ethical dimensions that may necessitate adjustments to the experimental design or procedures.

Furthermore, obtaining official ethical approval is beneficial for conducting school experiments. Schools may require ethical approval to grant access, an increasing number of journals demand it for publication, and approval can protect researchers when their study is ethically contentious. It is important for researchers to anticipate that the response to their ethical approval application may take several weeks. The outcome can vary, ranging from direct approval without alterations to conditional approval with required changes or outright rejection. Allocating sufficient time for implementing necessary modifications is advisable. At times, this may involve modifying the study design, a complex process that requires thoughtful consideration to ensure the continued appropriateness of the design for testing the study's hypotheses.

Obtaining consent in school experiments depends on the risks and benefits associated with the participants' involvement. Risks can range from mild discomfort, such as being presented with a challenging task or being alone with an unfamiliar research assistant, to more significant emotional distress, like sadness or anxiety triggered by the study. Upholding the principle of proportionality is crucial, ensuring that participants are not burdened unnecessarily beyond what is required to achieve the research objectives. For economists conducting school experiments, the risks are often similar to those encountered in regular school lessons, including potential cognitive load. These considerations should be taken into account when designing the study. To ensure that children are not overwhelmed and to accommodate their limited attention span, it is advisable to plan and execute the study in an enjoyable and seamless manner. Emphasizing the benefits of participation is also important, such as the opportunity to earn compensation, acquire knowledge, experience enjoyment, and foster class cohesion and exchange through a shared experience.

School experiments are conducted within educational institutions, involving various stakeholders such as teachers, headmasters, school administrators, research assistants, funding agencies, policymakers, and ministries. As stakeholders, these individuals and entities are subject to ethical considerations in the context of a school experiment. Therefore, researchers have a responsibility to take into account the interests of these parties throughout the study's design, implementation, and dissemination of results. This entails establishing regular communication channels to provide timely and relevant information before, during, and after the study. Obtaining consent from key stakeholders can also help facilitate the smooth implementation of the study.

Researchers bear the responsibility of upholding ethical principles in their school experiments. In recent years, the economics community has established institutions to foster ethical conduct in research. The practice of pre-registering studies has gained popularity, serving as a tool to meticulously plan and maintain study protocols while enhancing the credibility of the study's findings. Pre-registration can also be a publication requirement, with platforms like the AEA RCT Registry offering researchers a means to pre-register their studies.[12]

Balancing ethical and research standards can sometimes present conflicting priorities. On the one hand, researchers have the responsibility to consider ethical standards in the entire course of their school experiment. On the other hand, researchers have limited resources for handling ethical conduct. In certain cases, compliance with ethical standards may even clash with adherence to research standards. For instance, obtaining informed consent may raise concerns regarding potential selection bias in the collected data. Some parents or legal guardians may not receive the consent form from their children, while others may forget to return it. Additionally, language barriers can hinder understanding of the consent form. Therefore, researchers must ensure that the consent form is easily comprehensible, be available to address any queries, provide teachers with answers to commonly asked questions, and establish a supportive environment for all parties involved. To mitigate these conflicting priorities, researchers must allocate the necessary resources for the ethical conduct of the study during the planning phase of the school experiment. This includes dedicating time for reflection on ethical issues, obtaining Institutional Review Board (IRB) approval, and maintaining effective communication with stakeholders. By proactively addressing these concerns, researchers can navigate the challenges that arise and uphold both ethical and research standards.

2.2.5 Experimental procedures

The experimental procedures encompass a step-by-step protocol for the school visits and the overall implementation of the school experiment. The written protocol that is carefully followed by the research staff is important to ensure high data quality and replicability. It is important to recognize that the physical development, cognitive abilities, social growth, concentration, and experiences of children can vary based on factors such as age, sociodemographic characteristics, and the type of school they attend (Dahl et al., 2018; List et al., 2023). For example, children generally have shorter attention spans than adults, a greater inclination to move, and can be more easily emotionally triggered (Dahl et al., 2018; Punch, 2002). These characteristics should be considered when formulating the experimental procedures.

Similarly to the experimental design, the researchers' procedures should be clear and easy to comprehend. It is crucial to acknowledge that teachers and headmasters are often overwhelmed with their responsibilities and may not be familiar with the scientific norms and best practices of school experiments. To accommodate the needs of teachers, the procedures should minimize the time and responsibility required from them to comply with the protocol. As the research team possesses the expertise on how to conduct the experiment and establish a controlled environment, they should primarily take charge of implementing the protocol when conducting the experiment. This approach contributes to the quality of the data, as high compliance with the protocol enhances the consistency and control of the experiment. However, for large-scale intervention studies or program evaluations, it may not be feasible or advisable to have research assistants solely run the experiment.

[12] See e.g., `https://www.socialscienceregistry.org`

The potential for findings from the experiment to be scaled up is higher when teachers are involved in implementing the program, as the experimental conditions are more akin to the actual conditions when extrapolating the program (Al-Ubaydli et al., 2017). Ideally, teachers should be blind to the treatment during the experiment to mitigate demand effects and other threats to the study's internal validity. However, the feasibility and justification for maintaining teacher blindness depend on the specific context.

The procedures should establish an appropriate classroom environment conducive to a controlled experiment. This entails providing a quiet space with minimal distractions for the children, allowing them the freedom to make their own decisions. Depending on the nature of the experimental tasks, it may be necessary to allocate an additional room where tasks can be carried out with a small group of children at a time. Proper preparation of this room is crucial. Alternatively, if the experiment is conducted within the regular classroom setting with all students present, dividers on the tables and the use of tablets can ensure private answers and decision-making, unless spillover effects or interactions are intentionally desired based on the experimental design.

To maximize children's engagement with the experimental tasks and minimize restlessness, it is important to develop a protocol that enables children to maintain attentiveness throughout the session. To achieve this, the experiment itself should be inherently interesting (as discussed in the experimental design section). Additionally, the duration of sessions should be relatively short, and breaks can be included to allow children to recharge. It is worth noting that the attention span of children varies not only by age but also due to other factors such as physical activity levels (De Greeff et al., 2018) or the prevalence of ADHD (Barkley, 1997).

The protocol should also address the activities that take place during breaks between school lessons. It should specify whether children are allowed to interact with each other or if they are expected to continue working independently. This consideration is important as children's behavior during the experiment could be influenced by their experiences during these breaks.

Furthermore, the protocol should be tailored to the specific target group of the study. If the participating children belong to different grade levels, the protocol should be adjusted accordingly to accommodate their age and maturity levels. For example, sessions for younger children may need to be shorter and/or include more frequent breaks compared to sessions for older children.

The protocol should encompass provisions for addressing various scenarios that may occur during the classroom experiment. For instance, it should account for the emotional states of the participating children throughout the experiment, taking into consideration situations where some children may feel sad or disappointed if they receive fewer presents compared to their peers at the end of the experiment. Another common occurrence is when children are unable to participate due to not having obtained informed consent from their guardians. Additionally, in experiments involving multiple classroom visits, it is possible that some children may be absent due to illness during one or more of the visits. It is crucial for the research team to establish predefined procedures to effectively handle these recurring situations.

2.2.6 Timing of the experiment

Timing plays a crucial role in conducting a school experiment, and researchers must consider it in two key aspects. Firstly, the timing of the experiment within the school year needs careful attention. This entails checking vacation periods, holidays, and school schedules to effectively plan the study timeline. Researchers should also account for any potential time constraints that may arise from external factors, such as school closures during remote learning due to a pandemic. In cases where multiple schools are involved, researchers should initiate early communication with the schools to agree upon suitable dates and a precise schedule for conducting the study. As school schedules can sometimes change abruptly, it is advisable to contact the schools a day or a few days in advance to obtain a final confirmation of the exact schedule for the school visit.

Prior to contacting the schools, it is recommended to establish an implementation schedule that allows for sufficient buffer time for travel between school or classroom locations, equipment set-up, and removal. Additionally, researchers should allocate ample time between sessions and incorporate extra time to accommodate unforeseen challenges, such as technical issues with tablets or potential delays in the arrival of children. School and class activities, such as project weeks, excursions, or exam periods, may result in busy schedules, making it difficult to find suitable time slots. Therefore, researchers should be prepared for the possibility of encountering unpopular time slots that do not align with any school's schedule.

Secondly, it is advisable to keep the study duration as concise as possible to ensure a smooth implementation, considering the often tight school schedules. However, this can be a trade-off, particularly when evaluating interventions that require sufficient time with the children in the classroom to observe their effects. The breaks between lessons should also be taken into account, as they can disrupt the study's flow. Planning breaks in advance by discussing with teachers and students, potentially replacing the usual break with a longer one after the study, can help maintain participants' focus and engagement.

Timing considerations can have implications for the study results. Different implementation times can impact data quality, as factors like fatigue after lunch or variations in attention spans throughout the day may influence children's behavior, performance, and decision-making in the study. For example, children treated just before a holiday may respond differently compared to those treated immediately after a holiday. Moreover, timing can also impact the study's design, potentially leading researchers to modify certain features of the initial design due to time constraints in schools. Researchers may also consider incorporating timing into the randomization procedure. Stratifying the treatment classes/schools based on visit timing or documenting the visit times and controlling for them in the analysis can be helpful.

2.2.7 Support from partners

When conducting school experiments, the primary collaborators are likely to be teachers, school administrators, and public administrators. Regardless of whether researchers aim to conduct a lab-in-the-field experiment or implement and evaluate a comprehensive intervention, the support of these partners is crucial for the study's success for several reasons.[13]

Firstly, partners must grant permission to enter the schools and carry out the study. It is important to ascertain which authority has the jurisdiction to provide this permission, whether it lies with the schools autonomously or requires approval from a higher public authority. Additionally, parental or legal guardian consent will likely be necessary. The support of school administrators and teachers can be instrumental in obtaining such consent, leveraging the existing trust relationship between schools and parents/legal guardians. Schools possess the knowledge of the most effective methods to contact parents/legal guardians and secure their consent.

Secondly, schools provide essential resources, including classrooms, which are indispensable for conducting the study. Maintaining a positive rapport with the schools is vital for ensuring a smooth data collection process. For instance, establishing a good relationship with teachers can help guarantee the presence of all participating children in the classroom at the agreed-upon time, thereby enhancing the organization of data collection. Furthermore, teachers can assist in providing the necessary infrastructure, ensuring that children have access to computers, and preventing any conflicting activities from disrupting the implementation.

Thirdly, schools possess valuable insights into their pupils, such as their abilities and attention spans. Therefore, school administrators and teachers can provide valuable input on the appropriateness of the experiment for different age groups. They can offer feedback on the comprehensibility of instructions and the suitability of incentives for the participants. For example, researchers may need to decide whether to offer financial or material rewards, such as toys. In the case of offering toys, it is crucial to select items that children are fond of (see also Section 2.2.2).

Lastly, partners may provide valuable feedback that can have intriguing implications for researchers. This feedback may include information on potential issues when scaling up the intervention, a layperson's perspective on the study's results, or ideas for future follow-up studies. Moreover, engaged partners can actively promote the study and increase its outreach within the educational community.

Furthermore, the involvement of school administrators and teachers is crucial in both the implementation and evaluation of the intervention, as they typically assume responsibility for its execution. The extent of teachers' involvement may vary depending on the nature of the intervention being tested. Some interventions require teachers to directly deliver the content to the students, while others assign them a role primarily focused on encouraging student participation in the intervention. To enhance the interpretation of the study results, it is advisable to collect measures of compliance or implementation fidelity to assess the extent to which the treatment was implemented as intended (Dusenbury et al., 2003).

Schools can also sometimes contribute administrative data that can be used to control for background characteristics and observe outcome variables of interest. This is particularly advantageous when studying young children, as obtaining accurate information on their background characteristics through

[13]See also Glennerster (2017) for a discussion on the importance of partners and how to establish a good relationship with stakeholders in experiments more generally.

surveys can be challenging. Additionally, administrative data can provide real choices (e.g., school track selection) or performance-based outcome measures (e.g., grades) that are less dependent on children being present on specific data collection days, thereby minimizing missing data. Moreover, administrative data is less susceptible to measurement error compared to survey-based or experimental approaches.

How to establish a good relationship with the stakeholders? To achieve this objective, it is crucial to have a comprehensive understanding of the stakeholders' priorities. Educators are primarily focused on providing quality education to children. Therefore, effective and clear communication is essential to ensure that study participants and their parents or legal guardians comprehend the benefits of the intervention, which is critical for the successful implementation of the study. It is important to articulate the advantages of the study for the children and be flexible in making adaptations to maximize the efficient use of valuable instructional time. When teachers are convinced of the benefits that the study offers to the children, they are more likely to be motivated and supportive during the implementation phase. Furthermore, teachers operate within tight schedules and must cover a substantial amount of curriculum within a limited timeframe. Therefore, it is imperative to respect their time constraints and coordinate the implementation process in collaboration with them and the school administrators.

Additionally, maintaining a strong relationship with participants and stakeholders entails providing them with a debriefing and sharing the study results (e.g., through a dedicated website) at the conclusion of the research project.[14]

To solidify the collaboration and delineate responsibilities, it can be beneficial to establish a written cooperation agreement with the schools, clearly outlining the respective roles and obligations of the schools and the researchers. Additionally, when administrative data is collected, a data agreement is necessary to ensure proper handling and confidentiality of the data. J-PAL offers a comprehensive overview of the different types of agreements on its website.[15] The specific agreements required will depend on the legal requirements of the country in which the study is conducted.

3 Junior Survey

In the previous section, we have described the most important aspects when running school experiments based on an expert survey (see Subsection 2.1). In this section, we examine data obtained from a survey of junior researchers who are relatively new to the field of school experiments, primarily consisting of Ph.D. students or early-stage post-doctoral researchers. The junior survey serves two main purposes. Firstly, it assesses juniors' second-order beliefs about seniors' opinions on implementing field experiments in schools. Understanding these beliefs is vital for meeting academic requirements, increasing publication success, and conducting successful school experiments. Secondly, the survey gathers juniors' opinions on their ranking of school experiment's most important aspects, revealing differences

[14]Debriefing of participants and stakeholders should include descriptive results. Researchers should not use jargon and complex graphs but convey the results clearly and interestingly to the target group. For decision-makers/policymakers, researchers should provide useful and actionable results/implications of their research, if warranted by the study design and findings.

[15]See `https://www.povertyactionlab.org/resource/formalize-research-partnership-and-establish-roles-`

from seniors.

3.1 Recruitment and sample description of the junior survey

We employed the ”Economic Science Association” (ESA) mailing list to contact potential junior researchers for our survey. This list is a hub for scholars engaged in experimental social science research, offering us direct access to our target group – junior researchers with a keen interest in conducting school experiments. To encourage participation, we offered compensation of $50 to five randomly chosen participants as well as additional incentives depending on the treatment group.

We conducted the survey in October/November 2022. Our sample consists of 49 junior researchers who completed our survey. Nearly all of the researchers in our sample (98%) conduct their research in the field of economics, and most of them work at universities (89.8%). Upon closer examination of our demographic data, it becomes apparent that, overall, we have likely succeeded in primarily surveying junior researchers. Specifically, we find that the majority (76%) of the respondents stated that they are either Ph.D. students or early Post-Docs (Ph.D.: 49%, Post-Docs: 26.5%, 24.5% indicated “other”). Moreover, 41% of the respondents have not conducted any school experiment. Out of the group of respondents that have conducted a school experiment, the majority (59%) has not published more than one paper in total (36% published no paper, 23% published one paper) demonstrating large differences in experiences between juniors and seniors regarding the number of publications overall and experiences with planning and implementing school experiments.

3.2 Experimental Design and Results

In our junior survey, we presented the aspects identified from the expert surveys to junior researchers, who ranked them according to their believed importance for running effective school experiments. More precisely, in one treatment condition (“2nd Order Belief” condition, N=26), we asked juniors to assess which three aspects were most frequently mentioned by the 25 senior researchers. We do so to learn more about the junior researchers’ ability to anticipate seniors’ opinions. Second, to learn more about juniors’ own opinion, we asked juniors about their own views regarding the three most important aspects in the second treatment condition (“Own Opinion” condition, N=23). We incentivized second-order beliefs by informing participants when explaining the task that five randomly selected participants who accurately anticipated the aspects most often mentioned by seniors receive a $25 prize. Moreover, an additional prize of $25 was awarded if the ranking was also accurate.[16] The incentive for undertaking this task was supplementary to the compensation offered to five randomly selected participants, announced in the junior survey invitation.

Figure 1 provides a comparison of juniors’ second-order beliefs (gray bars) and their own opinions (white bars) to the percentage of seniors mentioning a specific aspect (black bars). For the junior data, we focus on the percentage of how often an aspect was ranked by the juniors among the top three,

[16] We had more than five participants who chose the correct aspects, but none of them also estimated the ranking correctly.

independent of the specific ranking. The figure presents the aspects in descending order from left to right, based on their frequency of mentions by the senior researchers.

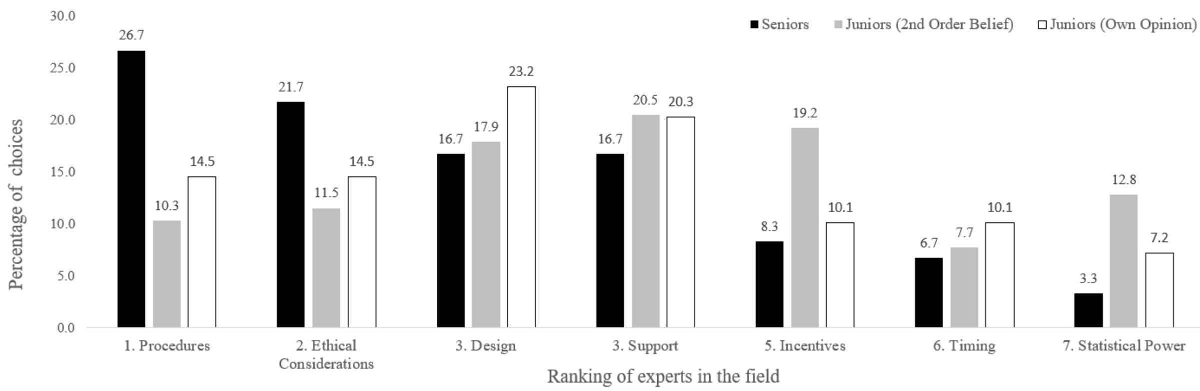

Figure 1: The aspects rank from one to seven according to the perceived importance by experts in the field. The figure shows the share of junior's total choices for each aspect and compares seniors' opinions (black bars) to the percentage of junior researchers' second-order beliefs (gray bars), and the percentage of junior researchers' own opinions (white bars) across aspects.

The figure underscores a discrepancy between seniors' and juniors' perceptions regarding the importance of certain aspects. Seniors prioritize "(Experimental) Procedures" and "(Ethical) Considerations," while juniors anticipate a different ranking, with "Support (from Partners)" and "(age-appropriate) Incentives" taking the top two positions.

Turning to juniors' own opinions, we find again that the junior's own priorities do not align with the seniors' top three mentioned aspects. That is, juniors' two most often chosen aspects for their top three ranking are "(age-appropriate) Design" and "Support (from partners)," in contrast to "(age-appropriate) Procedures" and "Ethical Considerations," which were most important to the experts. The latter aspects are chosen by about 15% of juniors in their top three ranking.

In summary, the survey highlights juniors' challenges in accurately predicting seniors' opinions and their differing priorities when conducting school experiments compared to experts. The junior's own priorities may either reflect the priorities of the future generation of school experimenters or indicate misjudgment due to a lack of experience in designing and implementing school experiments. While we acknowledge that the comparisons between seniors' and juniors' priorities in this chapter may not be entirely comparable due to measurement differences, we believe the junior survey serves as a proxy for priorities, highlighting the need for increased knowledge exchange between senior and junior researchers, as advocated in this guide.

4 Nine practical lessons for navigating school experiments and stakeholder relationships

The practical implementation poses a significant challenge for school experiments. Experts in the field provided insights on their top three lessons learned and strategies for fostering lasting stakeholder relationships in the realm of our senior survey (see Section 2). We categorized their responses into nine key lessons, which will aid junior researchers in navigating potential pitfalls and overcoming challenges in this area.

1. **Always take the peculiarities of your participants into account**

Throughout each phase of your study and every decision you make, it is crucial to consider the unique characteristics of your (desired) subject pool. Please note that depending on your recruitment strategy, the school system in your country, and the specific regions/districts of your recruited schools, you will have a selection of children with specific characteristics that may not be representative. You must develop a thorough understanding of their skills, needs, and environment. It is important to acknowledge the substantial heterogeneity that may exist in terms of children's cognitive and social development, as well as their sociodemographic background, both within and across classrooms.

Considering the unique characteristics of your subject pool should inform the research questions and hypotheses. This consideration should also impact various aspects of the study, including the study design (see subsection 2.2.1), procedures (see subsection 2.2.1), incentives (see subsection 2.2.2), recruitment, informed consent procedures (see also subsection 2.2.4), and the debriefing process of your school experiment.

2. **Develop a stakeholder engagement strategy**

It is important to conduct the experiments as planned without any major interruptions and obstacles, particularly during your experiment's implementation phase. To achieve this, you must establish a solid stakeholder engagement strategy and rely on the support of your stakeholders. Begin by identifying all the stakeholders who play a significant role in the successful conduct of your school experiment. The stakeholders can be closely affiliated with the schools, e.g., school authorities, headmasters, or external to the schools, such as other private or public institutions that have an interest in your study. Put yourself in the shoes of each stakeholder and understand their respective roles in your experiment, the potential benefits they can derive from it, their concerns, and their informational needs. When communicating with the stakeholders, use adequate information channels. Some stakeholders may prefer phone calls, while others may require written information materials. Prepare written materials tailored to different stakeholder groups in advance, so you are prepared when they request them. Regularly maintain communication with your stakeholders, and consider scheduling meetings or participating in existing platforms, such as parent or teacher meetings, to explain your work. It is essential to consider the perspective of stakeholders when communicating about your research. Emphasize your team's competence, the study's purpose, its societal value, and the specific benefits for the stakeholders to effectively convey the relevance and significance of your research. To do this successfully, researchers must understand the viewpoints of their stakeholders, which can be achieved through discussions with other researchers conducting school studies or by engaging directly with the stakeholders themselves. To ensure the effectiveness of your materials, you can seek feedback from colleagues by having them read and provide honest opinions. When engaging with stakeholders, treat them as equals and use language that is easily understandable, avoiding scientific jargon. Moreover, anticipate the stakeholder's concerns (e.g., "the teachers do not have the time to conduct your study") and address them either through design modifications (e.g., "our trained enumerators will conduct the study") or by providing convincing arguments (e.g., "the study is worth it because..."). See also subsection 2.2.7 for further information on securing support from partners.

3. **Build trust with stakeholders through transparency**

Researchers may feel inclined to withhold certain details of an experiment from stakeholders to avoid conflicts or to convince them to support the study. However, it is important to maintain transparency about the objectives, benefits, and also the risks associated with your school experiment.[17] Communicate a realistic timeline and provide an honest description of what you expect from your stakeholders and what you will provide to conduct the experiment. Supplying stakeholders with relevant and candid information and engaging them in discussions about the study will make them feel involved and invested. Take their feedback seriously and demonstrate a genuine commitment to their concerns. If you make promises or commitments to stakeholders, ensure that you fulfill them or promptly explain any changes to initial plans. By being forthright, you increase the likelihood of building lasting relationships with stakeholders and reduce the risk of conflicts, ultimately garnering their support during the critical implementation phase.

4. **Comply with ethical standards and data policies**

Make sure you comply with ethical standards and data policies, both in the country of your academic institution and the location where you conduct your study. Obtain the required approvals and written statements in accordance with the standards from authorized institutions. Identify any additional stakeholders from whom formal consent is necessary and ensure that you obtain their consent before commencing the implementation phase. As the principal investigator, you bear the responsibility for the proper execution of your study and complying with legal and ethical scientific requirements. It is crucial to meet these obligations to maintain the integrity of your research. For more detailed information on ethical considerations, see subsection 2.2.4.

[17]Note that the transparency requirements do not imply revealing treatment conditions to participants. When recruiting participants for experiments, it is often necessary to provide a general explanation of the research aims and study design to prevent demand effects between control and treatment groups. After the experiment is completed, stakeholders can receive a more detailed debriefing about the research design.

5. **Be aware of drop-outs and attrition and recruit sufficient participants**

It is important to anticipate relatively low participation and high drop-out rates in your school experiment. When determining your sample size, calculate with at least 10-20% fewer children than the total number of children in the participating classes. The actual participation rate can be influenced by various factors, such as the season (e.g., flu season), the type of study (e.g., multiple data collections for a panel study), the content of the study (e.g., a contentious subject that may make it difficult to convince all participants to take part), and country-specific aspects (e.g., generally low attendance rate in certain schools). Bear in mind that response rates from parents or legal guardians are expected to be even lower. If you intend to gather information on family background, consider including these questions in the children's survey or explore the use of available administrative data.

If your study involves multiple data collection points, it is essential to proactively address attrition and develop strategies to handle it during the data analysis phase. Additionally, consider the level of randomization (e.g., individual, class, or school level) when planning your recruitment to ensure an adequate sample size. For more detailed information on statistical power in school experiments, see also subsection 2.2.3.

6. **Build a reliable team for your implementation**

The successful implementation of your school experiment may require a team of data collectors or research assistants. It is crucial to have reliable, well-trained, and motivated individuals to ensure high-quality data collection. While school experiments share similarities with other data collections in this regard, you can find abundant online resources that provide detailed information on the hiring process, selection criteria, and training sessions for data collectors.[18]

However, school experiments differ from other data collections when it comes to the involvement of teachers, who may express interest in participating or assisting in the data collection process. While utilizing teachers as surveyors can potentially reduce costs, it also presents risks in terms of data quality, as there is a loss of control over adherence to the experiment's procedures. Teachers may have their own ideas or modify instructions, which can compromise the integrity of the data. If you decide to involve teachers in the data collection, it is crucial to provide them with adequate training beforehand to ensure consistency and adherence to the research protocol.

In general, teachers and other stakeholders play a vital role in the implementation of your study. It is crucial to secure their support and set aside time to explain the research objectives, clarify your expectations, and emphasize the benefits of their participation. For more details, we refer to Section 2.2.7.

7. **Bring essential equipment and consider the school's spaces**

[18] One such resource is available at this link: https://dimewiki.worldbank.org/Enumerator_Training.

It is important to consider the availability of necessary infrastructure within the participating schools. Beforehand, clarify with the schools which infrastructure they can provide and identify any additional equipment you may need to bring. For instance, if your study involves the use of electronic devices, it is crucial to ensure you have sufficient extension cords and multiple outlets, as classrooms may not have an abundance of outlets, and you may not always be situated near one. Keep in mind that Wi-Fi accessibility may vary, and it might be necessary to rely on local networks or set up your own Wi-Fi infrastructure, depending on the country and the participating schools.

The physical space for conducting your school experiment is also a crucial consideration. In many cases, classrooms serve as the primary space due to the limited availability of larger areas. To enhance privacy during decision-making, you can utilize desk partitions, creating an atmosphere similar to an examination setting and reducing noise interference. Furthermore, in smaller spaces, the use of tablet computers instead of pen and paper can be advantageous. Participants' attention is typically heightened, and using tablets allows them to privately log their answers and swiftly move on to the next screen.

8. **Plan ample time**

Coordinating a school experiment requires a significant amount of time, particularly when compared to a laboratory experiment. It is crucial to allocate an ample amount of time for the planning, coordination, and execution of your school experiment. Keep in mind that obtaining approval from school officials alone can take months.

When developing the timeline for your school experiment, carefully outline all necessary tasks, assign responsibilities, and set realistic deadlines. It is essential to communicate expectations and responsibilities clearly and adhere to the established schedule (see also subsection 2.2.6). Whenever possible, it is advisable to handle most of the tasks involved in conducting the experiment yourself and delegate only those that are absolutely necessary to the schools. This approach acknowledges that schools are typically occupied with their curriculum and the organization of their daily activities. Moreover, retaining as much control as possible over the correct execution of the experiment is desirable.

To adequately prepare for the implementation of your project, it is essential to plan ahead and address the items listed in the practical lessons (items 1 to 7). By anticipating potential challenges and difficulties in advance, such as time constraints that may arise despite initial support from schools, you can better cope with these issues. Allocating spare time specifically to accommodate unforeseen difficulties will provide you with the necessary flexibility and resilience throughout the project.

9. **Manage your risks**

School experiments inherently involve certain risks, which can manifest as deviations from the intended objectives of the study. These deviations can occur in various aspects, such as the scope, costs, and timing of the school experiment. The severity of a risk depends on both its probability of occurrence and the potential impact it may have on the study (for more information on risk assessments see, e.g., Ostrom and Wilhelmsen, 2019). For instance, a school may unexpectedly withdraw from participation, or you may encounter challenges in obtaining the desired outcome variable from authorities. To effectively manage these risks, it is crucial to identify and acknowledge them early in the study. Subsequently, it is necessary to define and implement risk mitigation measures to minimize any potential damages should the risks materialize. Timely awareness of these risks enables the development of appropriate strategies to address them. In the example mentioned, an alternative outcome variable could be defined and collected to mitigate the risk of not obtaining the first-choice variable. By considering the points outlined in these practical lessons (points 1 to 8), such as building trust with all stakeholders, the overall risk of a school dropping out on short notice can be reduced.

5 Conclusion

School experiments increase in significance in economics, necessitating researchers' awareness of the distinct characteristics inherent in the school context. This paper serves as a practical guide, presenting seven identified aspects specific to school experiments derived from expert insights experienced in conducting experiments in schools. Among the identified aspects, "experimental procedures" and "ethical considerations" stand out as most important when working with children, a vulnerable group. This requires researchers to exercise extra caution in safeguarding participants' well-being, allowing ample time for ethical approval. Additionally, seniors stress the significance of accounting for children's developmental stage in the experimental design, procedures, and incentive selection. The support of key stakeholders, such as headmasters and teachers, proves vital in conducting school experiments, emphasizing the value of fostering positive relationships with them. Lastly, aligning the experiment's timing with the school year and day, alongside careful consideration of statistical power, emerge as important aspects for the successful design and implementation of school experiments.

The survey of inexperienced researchers in school experiments revealed a disparity in perceptions compared to seniors regarding the most crucial aspects of school experiments. This underscores the need for improved communication and knowledge sharing among researchers of different career stages. This guide contributes to this goal, e.g., by summarizing the expert's nine lessons learned, which can be particularly beneficial for novices in the field of school experiments.

Overall, this comprehensive guide equips the reader with an overview of important considerations when planning their experiments to avoid pitfalls from the outset of the school experiment to a smooth implementation. By considering these distinct aspects of school experiments (compared to other field experiments), researchers can enhance the data quality, implementation fidelity, and validity of their studies and plan resources for their school experiments more effectively.

References

Al-Ubaydli, O., M. S. Lee, J. A. List, and D. Suskind (2021). The science of using science: A new framework for understanding the threats to scaling evidence-based policies. In The Scale-Up Effect in Early Childhood and Public Policy, pp. 104–125. Routledge.

Al-Ubaydli, O., J. A. List, and D. L. Suskind (2017). What can we learn from experiments? Understanding the threats to the scalability of experimental results. American Economic Review 107(5), 282–286.

Alan, S., N. Baydar, T. Boneva, T. F. Crossley, and S. Ertac (2017). Transmission of risk preferences from mothers to daughters. Journal of Economic Behavior & Organization 134, 60–77.

Alan, S., C. Baysan, M. Gumren, and E. Kubilay (2021). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. Quarterly Journal of Economics 136(4), 2147–2194.

Alan, S., S. Ertac, and M. Gumren (2020). Cheating and incentives in a performance context: Evidence from a field experiment on children. Journal of Economic Behavior & Organization 179, 681–701.

Alfonso, A., P. Brañas-Garza, D. Jorrat, P. Lomas, B. Prissé, M. Vasco, and M. J. Vázquez-De Francisco (2023). The adventure of running experiments with teenagers. Journal of Behavioral and Experimental Economics 106, 102048.

Andreoni, J., A. Di Girolamo, J. A. List, C. Mackevicius, and A. Samek (2020). Risk preferences of children and adolescents in relation to gender, cognitive skills, soft skills, and executive functions. Journal of Economic Behavior & Organization 179, 729–742.

Angerer, S., D. Glätzle-Rützler, P. Lergetporer, and M. Sutter (2015). Donations, risk attitudes and time preferences: A study on altruism in primary school children. Journal of Economic Behavior & Organization 115, 67–74.

Athey, S. and G. Imbens (2017). Chapter 3 - the econometrics of randomized experimentsa. In A. V. Banerjee and E. Duflo (Eds.), Handbook of Field Experiments, Volume 1 of Handbook of Economic Field Experiments, pp. 73–140. North-Holland.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of adhd. Psychological Bulletin 121(1), 65.

Bauer, M., J. Chytilová, and B. Pertold-Gebicka (2014). Parental background and other-regarding preferences in children. Experimental Economics 17, 24–46.

Ben-Ner, A., J. A. List, L. Putterman, and A. Samek (2017). Learned generosity? an artefactual field experiment with parents and their children. Journal of Economic Behavior & Organization 143, 28–44.

Bettinger, E., S. Ludvigsen, M. Rege, I. F. Solli, and D. Yeager (2018). Increasing perseverance in math: Evidence from a field experiment in norway. Journal of Economic Behavior & Organization 146, 1–15.

Bindra, P. C., D. Glätzle-Rützler, and P. Lergetporer (2020). Discrimination at young age: Experimental evidence from preschool children. Journal of Economic Behavior & Organization 175, 55–70.

Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. Evaluation Review 19(5), 547–556.

Brocas, I. and J. D. Carrillo (2020). Introduction to special issue "understanding cognition and decision making by children." studying decision-making in children: Challenges and opportunities. Journal of Economic Behavior & Organization 179, 777–783.

Bruhn, M. and D. McKenzie (2009, October). In pursuit of balance: Randomization in practice in development field experiments. American Economic Journal: Applied Economics 1(4), 200–232.

Camerer, C., S. Issacharoff, G. Loewenstein, T. O'donoghue, and M. Rabin (2003). Regulation for conservatives: Behavioral economics and the case for" asymmetric paternalism". University of Pennsylvania Law Review 151(3), 1211–1254.

Cappelen, A., J. List, A. Samek, and B. Tungodden (2020). The effect of early-childhood education on social preferences. Journal of Political Economy 128(7), 2739–2758.

Carroll, C., M. Patterson, S. Wood, A. Booth, J. Rick, and S. Balain (2007). A conceptual framework for implementation fidelity. Implementation Science 2, 1–9.

Charness, G., R. Cobo-Reyes, E. Eyster, G. Katz, Á. Sánchez, and M. Sutter (2023). Improving children's food choices: Experimental evidence from the field. European Economic Review 159, 104562.

Charness, G., J. A. List, A. Rustichini, A. Samek, and J. van de Ven (2019). Theory of mind among disadvantaged children: Evidence from a field experiment. Journal of Economic Behavior & Organization 166, 174–194.

Charness, G., A. Samek, and J. van de Ven (2022). What is considered deception in experimental economics? Experimental Economics 25(2), 385–412.

Chater, N. and G. Loewenstein (2023). The rhetoric of reaction, extended. Behavioural Public Policy 7(3), 838–845.

Choshen-Hillel, S., Z. Lin, and A. Shaw (2020). Children weigh equity and efficiency in making allocation decisions: Evidence from the us, israel, and china. Journal of Economic Behavior & Organization 179, 702–714.

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. Routledge.

Dahl, R. E., N. B. Allen, L. Wilbrecht, and A. B. Suleiman (2018). Importance of investing in adolescence from a developmental science perspective. Nature 554(7693), 441–450.

De Greeff, J. W., R. J. Bosker, J. Oosterlaan, C. Visscher, and E. Hartman (2018). Effects of physical activity on executive functions, attention and academic performance in preadolescent children: a meta-analysis. Journal of Science and Medicine in Sport 21(5), 501–507.

Duflo, E., R. Glennerster, and M. Kremer (2007). Chapter 61 using randomization in development economics research: A toolkit. Volume 4 of Handbook of Development Economics, pp. 3895–3962. Elsevier.

Dufwenberg, M., G. W. Harrison, et al. (2008). Peter bohm: Father of field experiments. Experimental Economics 11(3), 213–220.

Dusenbury, L., R. Brannigan, M. Falco, and W. B. Hansen (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. Health education research 18(2), 237–256.

Eckel, C. and N. Candelo Londono (2021). How to Tame Lab-in-the-Field Experiments, pp. 79–102. UK: Cambridge University Press.

Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2022). The preference survey module: A validated instrument for measuring risk, time, and social preferences. Management Science 69(4), 1935–1950.

Fehr, E., D. Glätzle-Rützler, and M. Sutter (2013). The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. European Economic Review 64, 369–383.

Gelman, A. and J. Carlin (2014). Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. Perspectives on Psychological Science 9(6), 641–651.

Glennerster, R. (2017). Chapter 5 - the practicalities of running randomized evaluations: Partnerships, measurement, ethics, and transparency. In A. V. Banerjee and E. Duflo (Eds.), Handbook of Field Experiments, Volume 1 of Handbook of Economic Field Experiments, pp. 175–243. North-Holland.

Glennerster, R. and K. Takavarasha (2013). Running Randomized Evaluations: A Practical Guide. Princeton University Press.

Gneezy, U., S. Meier, and P. Rey-Biel (2011). When and why incentives (don't) work to modify behavior. Journal of Economic Perspectives 25(4), 191–210.

Gneezy, U. and A. Rustichini (2004). Gender and competition at a young age. American Economic Review 94(2), 377–381.

Grosch, K., S. Haeckl, and M. G. Kocher (2022). Closing the gender stem gap-a large-scale randomized-controlled trial in elementary schools. CESifo Working Paper 17-081/I.

Guerini, R., L. FitzGibbon, and G. Coricelli (2020). The role of agency in regret and relief in 3-to 10-year-old children. Journal of Economic Behavior & Organization 179, 797–806.

Gummerum, M., Y. Hanoch, M. Keller, K. Parsons, and A. Hummel (2010). Preschoolers' allocations in the dictator game: The role of moral emotions. Journal of Economic Psychology 31(1), 25–34.

Harbaugh, W. T. and K. Krause (2000). Children's altruism in public good and dictator experiments. Economic Inquiry 38(1), 95–109.

Harbaugh, W. T., K. Krause, and L. Vesterlund (2002). Risk attitudes of children and adults: Choices over small and large probability gains and losses. Experimental Economics 5, 53–84.

Hausman, D. M. and M. S. McPherson (1993). Taking ethics seriously: economics and contemporary moral philosophy. Journal of Economic Literature 31(2), 671–731.

Hermes, H., F. Hett, M. Mechtel, F. Schmidt, D. Schunk, and V. Wagner (2020). Do children cooperate conditionally? adapting the strategy method for first-graders. Journal of Economic Behavior & Organization 179, 638–652.

Kaplan, L., J. Kuhnt, and J. I. Steinert (2020). Do no harm? field research in the global south: Ethical challenges faced by research staff. World Development 127, 104810.

Kosse, F., T. Deckers, P. Pinger, H. Schildberg-Hörisch, and A. Falk (2020). The formation of prosociality: causal evidence on the role of social environment. Journal of Political Economy 128(2), 434–467.

List, J. A. (2008). Informed consent in social science. Science 322(5902), 672.

List, J. A. (2011). Why economists should conduct field experiments and 14 tips for pulling one off. Journal of Economic Perspectives 25(3), 3–16.

List, J. A., R. Petrie, and A. Samek (2023). How experiments with children inform economics. Journal of Economic Literature 61(2), 504–564.

McConnell, B. and M. Vera-Hernández (2022). More powerful cluster randomized control trials. Technical report, IFS Working Paper.

Ostrom, L. T. and C. A. Wilhelmsen (2019). Risk assessment: tools, techniques, and their applications. John Wiley & Sons.

Ozturk, O. D., E. A. Frongillo, C. E. Blake, M. M. McInnes, and G. Turner-McGrievy (2020). Before the lunch line: effectiveness of behavioral economic interventions for pre-commitment on elementary school children's food choices. Journal of Economic Behavior & Organization 176, 597–618.

Piovesan, M. and H. Willadsen (2021). Risk preferences and personality traits in children and adolescents. Journal of Economic Behavior & Organization 186, 523–532.

Punch, S. (2002). Research with children: The same or different from research with adults? Childhood 9(3), 321–341.

Raudenbush, S. W., A. Martinez, and J. Spybrook (2007). Strategies for improving precision in group-randomized experiments. Educational Evaluation and Policy Analysis 29(1), 5–29.

Resnik, D. B. (2018). The ethics of research with human subjects: Protecting people, advancing science, promoting trust, Volume 74. Springer.

Rubinstein, A. (2001). A theorist's view of experiments. European Economic Review 45(4-6), 615–628.

Schochet, P. Z. (2008). Statistical power for random assignment evaluations of education programs. Journal of Educational and Behavioral Statistics 33(1), 62–87.

Schunk, D. and I. Zipperle (2023). Fairness and inequality acceptance in children and adolescents: A survey on behaviors in economic experiments. Journal of Economic Surveys 37(5), 1715–1742.

Sunstein, C. R. (2016). The ethics of influence: Government in the age of behavioral science. Cambridge University Press.

Sutter, M. and D. Glätzle-Rützler (2015). Gender differences in the willingness to compete emerge early in life and persist. Management Science 61(10), 2339–2354.

Sutter, M. and A. Untertrifaller (2020). Children's heterogeneity in cooperation and parental background: An experimental study. Journal of Economic Behavior & Organization 171, 286–296.

Sutter, M., L. Yilmaz, and M. Oberauer (2015). Delay of gratification and the role of defaults—an experiment with kindergarten children. Economics Letters 137, 21–24.

Sutter, M., C. Zoller, and D. Glätzle-Rützler (2019). Economic behavior of children and adolescents – a first survey of experimental economics results. European Economic Review 111, 98–121.

Ten Have, H. (2015). Respect for human vulnerability: the emergence of a new principle in bioethics. Journal of Bioethical Inquiry 12(3), 395–408.

Thaler, R. H. and C. R. Sunstein (2003). Libertarian paternalism. American Economic Review 93(2), 175–179.

Tobol, Y. and G. Yaniv (2019). Parents' marital status, psychological counseling and dishonest kindergarten children: An experimental study. Journal of Economic Behavior & Organization 167, 33–38.

Voslinsky, A. and O. H. Azar (2021). Incentives in experimental economics. Journal of Behavioral and Experimental Economics 93, 101706.