The following example is a full proposal submitted by Ariel Kalil and the BIP Lab (University of Chicago), shared here to illustrate how a strong application can meet J-PAL's evaluation criteria. (Note that this example is not specific to a specific sector/initiative area.)

I. Abstract

We propose a pilot study of Chat2Learn-AI, an AI-enhanced digital intervention for low-income parents to help them foster their preschool and kindergarten age children's language skill and curiosity. Chat2Learn-AI provides parents with conversation prompts for open-ended parent-child conversation. The pilot study will address three key objectives: 1) the feasibility and efficacy of Chat2Learn-AI for boosting the quantity and quality of parent-child language use, 2) the impact of Chat2Learn-AI on parental agency and growth mindset, and 3) the feasibility of recruiting and retaining low-income parents in an opt-out model through school partnerships. This pilot will inform the design and implementation of a full-scale RCT in 2026 that will evaluate the impact of Chat2Learn-AI on low-income young children's language skill and curiosity.

Proposals should begin with a brief abstract of 150-200 words.

Pilot proposals should clearly outline the expected learnings from the pilot and how they it will inform a full-scale RCT.

II. Background

J-PAL NA funds randomized evaluations that contribute to our mission of reducing poverty by ensuring policy is informed by rigorous evidence. Research questions should address a pressing public policy issue with significant impact on people who are low-income or living in poverty, who have risk factors associated with falling into poverty, or who identify as members of a racial/ethnic group at greater risk of living/falling into poverty due to structural racism. Proposals should discuss the policy landscape, potential contributions of the proposed evaluation to public policy, and potential to generate benefits for economically marginalized populations.

Many children enter kindergarten without the school readiness skills they need to thrive. Educators define school readiness both in terms of a child's language and math skills and with such qualities as curiosity and enthusiasm for learning (Snow & Matthews, 2016; ELOF, 2022). Compared to their more advantaged peers, children from low-income households enter kindergarten behind on school readiness skills (Kalil & Ryan, 2020). This gap reflects differences in parents' engagement in children's learning inside the home as well as differences in children's out-of-home education and care experiences (Mayer et al., 2019).

Most policy discussions about boosting low-income children's school readiness skills focus on changing their out-of-home experiences. Early childhood education can boost low-income children's school readiness skills in the short run but evidence for long-run impacts is inconsistent (Burchinal et al., 2024; Gallegos & Garcia, 2024). Income-based gaps in children's school readiness skills are unlikely to narrow without changing parents' engagement in children's learning (Mayer et al., 2019). To do so effectively we developed Chat2Learn-AI.

Chat2Learn-AI aims to build children's vocabulary and curiosity by improving caregiver-child oral language interactions in the home. Children's language development is fostered through listening to adult-driven oral language (Rowe, 2008) and children who hear more words spoken to them know more words. Back-and-forth talk (as opposed to overheard speech) is not only associated with linguistic growth but is positively correlated with children's cognitive, social, and emotional development (Golinkoff et al., 2019). Building vocabulary early in life is important for long-term learning (Cunha & Heckman, 2007) and improvements in early-childhood vocabulary are correlated with non-cognitive skills including self-control and executive function (Golinkoff et al., 2019). Children from lower income households are exposed to less language in the home and their parents are less likely to initiate and sustain them in conversation (Golinkoff et al., 2019; Rowe, 2008). Lower income children in turn have smaller vocabularies and slower vocabulary growth in early childhood and as many as 65% of children from low-income backgrounds exhibit clinically significant language impairment (Roberts et al., 2019). Boosting interactive caregiver-child language interaction is thus key to narrowing gaps in children's school readiness.

Proposals should situate their work within the existing literature to demonstrate how the study contributes to advancing knowledge. Strong proposals clearly compare their approach to prior research—including work outside of economics when relevant—to show the added value of the proposed evaluation.

The following example is a full proposal submitted by Ariel Kalil and the BIP Lab (University of Chicago), shared here to illustrate how a strong application can meet J-PAL's evaluation criteria. (Note that this example is not specific to a specific sector/initiative area.)

J-PAL defines pilots very broadly; pilots may conduct a small randomized evaluation to test aspects of design or implementation before scaling up, focus on planning activities only, or fall somewhere in between. While randomization is not required at the pilot stage, pilot proposals must clearly state how work will lead to a future RCT.

Chat2Learn-AI builds directly from our recently completed J-PAL-funded intervention "Evaluating Learning Interactions" (ELI). ELI was a 6-month intervention for lower-income parents to help them boost their children's vocabulary and curiosity. ELI provided text prompts (accompanied by an illustration) to parents designed to elicit parent oral language exchanges with their child. ELI sent three prompts per week to a sample of ~ 600 parents in the United States. The study had two treatment arms and a control group. In the first treatment arm, parents were randomly assigned to receive prompts to orally define a word for their child and to check the child's understanding of the word's definition. In the second treatment arm, parents were prompted to orally define the word for their child plus ask the child an open-ended question using that word. Children's knowledge of intervention words increased in both treatment arms (effect sizes of .65 and .42 SD, respectively). Children's knowledge of non-intervention words increased in the first treatment arm for children at the higher end of the endline assessment distribution (effect size of .30 SD). The results also suggested that ELI boosted children's curiosity, although these increases in curiosity were not statistically significant at p=.05.

Many pilot studies do not build directly on prior interventions. However, if your proposal does, including information on preliminary results can strengthen your case and provide review boards with valuable context on the viability of the project.

ELI provides important causal evidence that parental language input boosts children's vocabulary in a light-touch modality. ELI prompted parents to talk more with their child which meant their children heard and learned more of the words their parents spoke to them. ELI had a high retention rate (98%) and parents reported that their children responded to the program enthusiastically. However, ELI sent the same static prompts to everyone. For the reasons we outline below we believe that Chat2Learn-AI can further increase disadvantaged children's vocabulary of words not included in the intervention and that it will have a greater effect on curiosity.

First, in ELI children mostly learned the words that we selected for the intervention. Only some children learned novel words that were not part of the intervention. Parents can in principle teach their children many new words by asking follow-up questions and expanding their conversations to include a broader range of vocabulary and ideas. Parents may benefit from additional support to sustain engaging conversations. Guidance to parents on how to ask follow-up questions tailored to the child's response could be especially effective in maintaining conversation.

Chat2Learn-AI improves upon the static ELI program by incorporating new advances in technology that support dynamic, personalized, and expansive oral language interactions (see also Doss et al., 2019). AI integrations enable an interactive experience between the program and the parent and child. Instead of providing a single static prompt, Chat2Learn-AI personalizes unlimited conversation prompts in real time based on children's interests. In the proposed pilot, we will test whether treated parents increase two-way, open-ended conversation using more diverse vocabulary.

Chat2Learn-AI also incorporates new insights from ELI into the role of parent beliefs in children's learning. Incomplete information or biased beliefs are thought to underlie differences between advantaged and disadvantaged parents' behavior (Rowe, 2008). ELI significantly

reduced parents' adherence to a fixed mindset (reducing by five percentage points from a baseline average of 18%) (Rury et al., 2024). It did so without shifting parents' feelings of stress, fatigue, or enjoyment of learning activities. This demonstrates that the experience of doing an action provides important feedback that creates a reinforcing mechanism for behavior change.

Boosting young children's curiosity is an important goal of Chat2Learn-AI (see Alan & Mumcu, 2024, for evidence of the causal effect of curiosity on children's learning in elementary school) and we will test Chat2Learn-AI treatment impacts on children's curiosity in the full-scale RCT in 2026. Language development is more than learning a string of words; it is also the expansion in thinking, new ideas and new concepts that come with learning new words. We expect that as parents engage in longer and deeper conversation guided by the child's interest, these conversations will prompt the child's interest in discovery, resulting in an increase in measured curiosity (Loewenstein, 1994). In this pilot test we will measure the number of open-ended questions parents ask of children. Open-ended questioning is known to benefit children's curiosity development.

III. Chat2Learn-AI Proof of Concept

Chat2Learn-AI was developed by our research team and aims to build vocabulary and curiosity for children in preschool and kindergarten. Chat2Learn-AI is intended for low-income families and enrolls parents directly. It provides 3-5 weekly conversation prompts by text paired with colorful illustrations to engage parents in dynamic, high quality language interactions with their young children. The program encourages parents to engage in back-and-forth, open-ended conversation with their children about a variety of topics of interest to the child. Chat2Learn-AI personalizes unlimited conversation prompts in real time based on children's interests.

While prior piloting is not required to apply, proposals should demonstrate viability; this could be shown through clear and realistic timelines, a discussion of the partnership with implementing organizations, discussion of similar studies, etc.

We conducted a small lab experiment (n=22) as an implementation test for Chat2Learn-AI. Parent-child pairs were randomized into a control or treatment condition. The control group was asked to wait in a room for 10 minutes with no further instruction. The treatment group was asked to wait in a room for 10 minutes and was invited to use Chat2Learn-AI, which was offered on a digital tablet. This design aimed to simulate real-world conditions of access to Chat2Learn-AI, which in practice can be used on a tablet or a smart phone. We recorded and coded parent-child interactions during the 10-minute session. Although the results are not statistically significant in this very small sample they are promising. Treated parents said 36% more words and 6% more sentences than parents in the control group. Additionally, parents in the treatment group used more diverse vocabulary and employed longer, more complex sentences—key elements in fostering children's language development. The Chat2Learn-AI condition is associated with a greater frequency of parent responses relative to child talk, suggesting more extensive parental verbal engagement. Note that the lab experiment was a one-time observation. We anticipate a stronger impact of Chat2Learn-AI when families use it repeatedly.

III. The Pilot

Once parents are recruited (see details below), we will obtain consent and randomly assign them to either a treatment group that receives Chat2Learn-AI or a business-as-usual control group that does not receive any program. We will collect basic demographic information on participating families from our school partners. These data will be used as controls in analyses, to perform balance tests, and to conduct subgroup analyses. We will ask parents to complete a very brief

endline survey to assess the treatment's impact on parental beliefs such as growth mindset and sense of agency and to better understand parents' and children's experience with the program. We will observe and evaluate parent-child oral language quality pre- and post-intervention via an audio-recording task (described further below). As in our proof-of-concept implementation test, we will code these data for quality and quantity of syntax, sentiment, linguistic diversity, and conversational turns. These will be our key outcome measures for this pilot.

Proposals should clearly outline the pilot activities, in addition to the long term goals of the RCT. If the pilot is testing an intervention, proposals should provide clear details on how the intervention will be set up, delivered, and measured. This information helps reviewers assess the value and feasibility of the proposed work, and alignment with the study's research questions.

The Chat2Learn-AI pilot will last 12 weeks as our prior J-PAL funded pilot PEAKS. During these 12 weeks, parents in the treatment group will receive three to five illustrated conversation prompts by text per week. Parents can engage with the program as frequently or infrequently as they choose. The control group will not receive any text-based program and will only receive texts related to implementing the pilot including texts about consent, audio-recording tasks, and the survey. Based on our previous experience we are confident in our ability to implement this approach effectively and in parents' receptiveness to such messages. We observed high program engagement (as measured by the treatment-control difference in the rate of parental text-back to prompts) in our short 12-week pilot for PEAKS. Similarly, for the Chat2Learn-AI pilot, we will assess parental engagement by measuring the frequency of parents' texting back to us children's responses to the prompts generated by the Chat2Learn-AI program.

Pilots are encouraged to submit a letter of support from the implementing parter(s), but this a requirement for full studies only.

Recruitment. We will partner with 1-2 elementary school networks to recruit 200 low-income parents of kindergarten children (both Spanish and English speaking). We have attached a letter of support from Distinctive Schools, a K-12 charter school network in Chicago. Distinctive Schools serves more than 350 kindergarten students, 85% of whom are from low-income families eligible for free and reduced lunch. The racial makeup of their kindergarten students is 54% black, 40% Hispanic, 2% white, 2% multiracial, and 2% other. Distinctive Schools and any other school partners will enter into a data use agreement (DUA) with our research team to enable an opt-out model of consent for this pilot study. With a DUA in place, we will obtain primary parent contact information and basic family demographic information for all participating kindergarten classrooms.

Parents will be informed about the study partnership in advance and can opt-out at any time before or after randomization activities. Parents will be offered $20 for completing each of the audio-recording tasks at baseline and endline and an additional $10 for completing the endline survey for a total compensation amount of $50 for completing all data collection tasks.

Power. The primary outcomes of this pilot are the quality and quantity of parent-child talk. For ease of interpretation, the power analysis is conducted considering the effects in standard deviation units. To calculate the minimum detectable effect size (MDES) given our sample size of 200, we use PowerUP![1]. Given a 0.05 significance level, 0.8 power, and an expected R-squared of 0.45, power estimates suggest an MDES of 0.295 standard deviations. Our expected R-

Pilot proposals should include a preliminary, "back-of-the-envelope" discussion of power and potential sample size to detect relevant impacts if a full randomized evaluation were launched.

[1] Dong, N. and Maynard, R. A. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs. Journal of Research on Educational Effectiveness, 6(1), 24-67. doi: 10.1080/19345747.2012.673143

squared is based on previous projects by the BIP Lab that also include baseline measures of the outcome in the regression. Given the large effects observed on child vocabulary for ELI, we find the MDES for this pilot to be reasonable.

The Chat2Learn-AI Intervention.

Chat2Learn-AI is powered by large language models (LLMs) that deliver dynamic conversation prompts to parents, tailored to children's interests as indicated by parent-recorded responses to prompts. Below is an example of an interaction with the program, and examples of prompts with accompanying illustrations can be viewed here: https://biplab.uchicago.edu/chat2learn/experiment.

> *Chatty-AI[2]: Ask Jamie – If you could choose to have any pet, which pet would you choose?*
> *Mom: Jamie would choose a horse.*
> *Chatty-AI: Awesome! Ask Jamie – If you could go on a trail ride with your horse, where would you go? A forest, beach, mountain, or somewhere else?*

Chat2Learn-AI delivers a foundation of three to five identical conversation prompts to all treatment families paired with illustrations each week. In addition, it layers on AI chatbot functionality, allowing the parent and child to continuously engage with the program on a countless range of topics that promote ongoing conversational engagement. Parents and children can spontaneously direct the program to generate prompts tailored to their specific interests to support longer, richer and more engaging conversations. Within the scope of the proposed pilot study, we will explore potential changes in various aspects of parent-child conversations following a period of using Chat2Learn-AI. Understanding these changes will provide critical insights into why Chat2Learn-AI might be effective and guide the design of a large-scale randomized controlled trial (RCT) to evaluate the program's impact on key outcomes, including vocabulary and curiosity.

We hypothesize that Chat2Learn-AI will affect parent behavior in conversational interactions with their child by increasing vocabulary, open-ended questions, and conversation turns and that it will increase parent growth mindset and parental agency for supporting their children's skill development.

As part of their discussion of academic relevance, proposals should note if research question(s) have previously been evaluated (including outside of economics) and draw clear links to underlying social science theories.

Outcomes.

<u>*Parent Engagement.*</u> Parent engagement will be systematically assessed throughout the intervention by tracking the frequency of responses to Chat2Learn-AI prompts. Specifically, we will track and analyze whether and how often parents text back their child's responses to the program prompts over the course of the intervention. Our JPAL-funded PEAKS pilot project successfully used this approach to test the impact of differential behavioral messages on parents' engagement with the program as a proxy for engagement with their child.

<u>*Parent Beliefs.*</u> In an endline survey, we will examine the treatment's impact on parental beliefs regarding (a) the malleability of their child's cognitive abilities (i.e., growth mindset), (b) the

[2] Chatty-AI is an abbreviation of "Chat2Learn-AI." This is the parent-facing name of the program.

value of engaging in rich language interactions with their child, and (c) their own sense of agency and interest in engaging in back-and-forth conversations with their children.

Parent-Child Talk. The primary outcome of interest for this pilot study is the quantity and quality of parent-child talk. To evaluate changes in parent-child conversations, parent-child pairs from both study conditions will participate in two time-constrained, audio-recorded conversational tasks pre- and post-intervention. Parent-child pairs will be invited to hold a conversation for five minutes without the facilitation of the Chat2Learn-AI program or any other conversation guide. These recordings will be transcribed and systematically coded to measure a) the overall number of sentences and conversational turns; b) the complexity of sentence structure and the diversity of vocabulary used; c) sentiment (i.e., positive and negative tone; praise and criticism) and d) the frequency of open-ended and follow-up questions posed by parents.

Pilot proposals should provide a clear timeline and discussion for how activities will inform or lead to a full randomized evaluation.

Timeline.

15 months. April – July 2025: Planning, development and piloting of intervention and audio task and recruitment of participants; *August - September 2025:* Administrative data and baseline audio task collection and randomization of participants; *October – December 2025:* Pilot implementation; *January – February 2025:* Endline audio task and survey collection; *March – June 2026:* Data cleaning, analysis, writing, and dissemination.

References

Proposals should include references, Chicago Author-Date style is suggested. References are not included in the page count.

Alan, S., & Mumcu, I. (2024). Nurturing childhood curiosity to enhance learning: Evidence from a randomized pedagogical intervention. *American Economic Review*, *114*(4), 1173-1210.

Burchinal, M., Whitaker, A., Jenkins, J., Bailey, D., Watts, T., Duncan, G., & Hart, E. (2024). Unsettled science on longer-run effects of early education. *Science, 384*(6695), 506-508.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American economic review*, *97*(2), 31-47.

Doss, C., Fahle, E. M., Loeb, S., & York, B. N. (2019). More than just a nudge: Supporting kindergarten parents with differentiated and personalized text messages. *Journal of Human Resources*, *54*(3), 567-603.

Effective Practice Guides: Head Start Early Learning Outcomes Framework (ELOF). (2022, June 10). U.S. Department of Health & Human Services, Administration for Children & Families, Office of Head Start, Early Childhood Learning & Knowledge Center. Retrieved August 6, 2022, from https://eclkc.ohs.acf.hhs.gov/school-readiness/effective-practice-guides/introduction

Gallegos, S., & García, J. L. (2024). Childcare and parenting in the production of early life skills. *Economics of Education Review*, *101*, 102557.

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child development*, *90*(3), 985-992.

Kalil, A., & Ryan, R. (2020). Parenting practices and socioeconomic gaps in childhood outcomes. *The future of children*, *30*(1), 29-54.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*, 75-98.

Mayer, S. E., Kalil, A., Oreopoulos, P., & Gallegos, S. (2019). Using behavioral insights to increase parental engagement: The parents and children together intervention. *Journal of Human Resources*, *54*(4), 900-925.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of child language*, *35*(1), 185-205.

Rury, D., Kalil, A., Mayer, S., & Bresciani, D. (2024). *Prompting parent-child talk boosts children's vocabulary and changes parents' beliefs about learning: Evidence from a field experiment.* Manuscript under review.

Snow, C. E., & Matthews, T. J. (2016). Reading and language in the early grades. *Future of Children*, *26*(2), 57–74.