Impact Evaluation of Bebbo

March 21, 2024

Introduction

Objectives of Bebbo

The two main objectives of Bebbo, in line with the UNICEF ECARO Early Childhood Development Theory of Change, are: (1) Improving availability of information for parents on child development, and (2) Supporting parents for responsive caregiving and early intervention. Accordingly, Bebbo provides users information and interactive tools to help nurture and aid their children's health and development. The launch of Bebbo in 11 countries in the ECA region is a direct response to the identified objective to engage parents and caregivers in nurturing care, positive parenting, and stimulating learning.

Bebbo Theory of Change(ToC)

If there are ECD services and systems in place to support implementation of Bebbo and

If Parents/caregivers receive sufficient information about Bebbo, and

If they are convinced to download the app, and

If they regularly access and use key functions of Bebbo

Then parents and caregivers will increase their knowledge and awareness of key aspects of child development; which would result in increased parental engagement in nurturing care, positive parenting practices, stimulating and learning activities.

Evaluation Design

The current multi-country experimental evaluation seeks to answer the following questions among a sample of Serbian and Bulgarian parents of 0-6 years old children:

1. Does being asked to use Bebbo improve parents' knowledge and awareness about child development and health, as well as their parenting confidence and attitudes?
2. Does being asked to use Bebbo improve positive parenting practices?

This is a randomized encouragement design. We recruited respondents on social media and measured effects across the three domains by administering survey questions four weeks after treatment and again eight weeks after treatment.

Evaluation Assumptions

If study participants are asked to download Bebbo

Then a significant portion will download Bebbo and regularly access and use its key functions and

Then they will increase in their knowledge and awareness of key aspects of child development; which would result in increased parental engagement in nurturing care, positive parenting practices, stimulating and learning activities.

Study Design: Measurement

How do we measure an *increase in their knowledge and awareness of key aspects of child development; which would result in increased parental engagement in nurturing care, positive parenting practices, stimulating and learning activities*?

Child Development Knowledge To assess knowledge and awareness on child development, parents were asked how to monitor the development of their child with 4 items (items: "I know how to monitor the language development of my child", "I know how to monitor the cognitive development of my child", "I know how to monitor the social-emotional development of my child", "I know how to monitor the physical development of my child").

Vaccine Knowledge For health awareness and knowledge, parents were asked on whether they know about the vaccination related needs of their children with the following item: "I know which vaccine my child needs to take next"

Parenting Confidence Parents were asked how confident they feel in their ability to deal with their child's emotions and how confident they feel in their ability to respond properly when their child misbehaves, on a 4-point Likert type scale (1=Not confident at all, 4=Very confident).

Attitude to Physical Punishment Attitude towards physical punishment was assessed with the following item: "Do you agree that in order to bring up, raise, or educate a child properly, the child needs to be physically punished?", on a 4-point Likert scale (1=Do not agree at all, 4=Strongly agree).

Activities Past 24h Parental engagement was measured using the parental engagement subscale of the UNICEF MICS Early Childhood Development Index module. Participants were asked to report if they were engaged in six parental activities such as reading books, telling stories, singing songs, taking the child outside, playing with the child and naming, counting, and drawing with the child in the last 24 hours.

Breastfed Breastfeeding practices of parents of 0-2 years old children were measured with the following yes/no item: "Has your child been breastfed in the last 24 hours?"

Measurement

Positive affect and hostile parenting practices were measured using the two relevant subscales of the Preschool Parenting Measure:

Positive Practices The positive affect subscale measures the warmth and affection involving parent-child interaction (4 items, example item: "I often smile when I am around my child")

Hostile Practices The hostility subscale measures harsh interactions between the parent and child (4 items, example item: "When my child does something wrong, I sometimes threaten her".

Experimental Design

Experimental Design

A randomized controlled encouragement design is conducted, in which study participants are randomized to one of the two following conditions:

Treatment. Participants in the treatment condition were told that there was one more step to qualify for the study and were then asked to download the app Bebbo and use it regularly, being encouraged that doing so will help them with their parenting.

Control. Participants in the control condition were told that there was one more step to qualify for the study and were then provided a link to a basic website containing information on parenting and asked to visit it regularly, being encouraged that doing so will help them with their parenting.

Experimental Design

The choice to use a website as a treatment-as-usual (TAU) condition was decided by the evaluation and program team because it represented an alternative (and traditional/existing) way to solve the problem that the Bebbo app was trying to solve. Another option that was considered was to use an alternate parenting app, but the team believed that using a website gave the best chance to detect a difference in the use of an app, rather than the specific implementation of the Bebbo app.

The downside with choosing a treatment-as-usual condition is that if one does not find significant effects of the treatment, one cannot differentiate between the following scenarios:

1. The control is effective and the treatment equally effective.
2. Neither the control nor the treatment are effective.

Study Design: Recruitment

Recruitment

Participants were recruited to the study with social media ads on the Meta platform (Facebook and Instagram) using the Virtual Lab platform to create and run the recruitment ads. The Virtual Lab platform is used to track and measure the price-per-respondent across multiple strata. It solves the core problem of monitoring, computing expectations, and adjusting budget when recruiting samples via social media platforms while ensuring they are representative across desired and measured characteristics.

In exchange for participating in the study, participants were told they could receive gift cards worth up to 12 USD (in their local currency).

Figure 1: Recruitment Ads

Recruitment

The survey was administered via a chatbot in Facebook Messenger, using the Virtual Lab platform. Respondents who clicked on the advertisements were directed to a Messenger chat.

The Virtual Lab chatbot allowed the researchers to create multi-wave surveys, with independent timing, with integrated gift card provision at the end of each wave. Gift cards were Visa International Cards administered by the Tremendous platform.

Recruitment

The inclusion criteria for sampling was to have a child between 0-6 years old.

The data collection was longitudinal in which we measured the outcomes of interest before treatment (in a baseline survey) and after treatment (in an endline survey).

Recruitment and survey administration was performed on a rolling basis between March and October, 2023. Each individual participant was treated at the end of the baseline survey, sent the endline survey 4 weeks after completing the baseline survey, and the follow up survey 4 weeks after the endline.
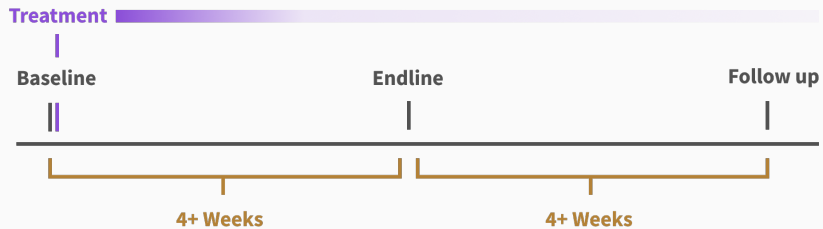
Figure 2: Study Design

Preliminary Analyses

Baseline Results

Table 1: Outcome Construct Descriptives Pooled Baseline

name	mean	median	min	max	sd	prop_max	prop_na
Activities Past 24h	5.05	5.00	0	6	1.21	0.48	0.0
Parenting Confidence	3.36	3.50	1	4	0.64	0.38	0.0
Positive Practices	3.10	3.25	1	4	0.80	0.26	0.0
Attitude to Phys. Punishment	3.08	3.00	1	4	0.89	0.37	0.0
Hostile Practices	3.08	3.00	1	4	0.69	0.17	0.0
Child Dev. Knowledge	0.87	1.00	0	1	0.27	0.75	0.0
Vaccine Knowledge	0.74	1.00	0	1	0.44	0.74	0.6
Breastfed	0.41	0.00	0	1	0.49	0.41	0.6

Baseline Results

Many of the constructs have quite high means and medians and some have a high proportion of respondents with the max score. In particular, 74% and 75% of respondents scored perfectly on the knowledge questions.

This is problematic, as knowledge is often considered the easiest to change quickly and was a core outcome of interest for the team. Additionally, knowledge questions seem to be heavily impacted by the repeated survey effect, as discussed further down.

Attrition

Table 2: Attrition: Pooled

stage	count	attrition	treated_attrition	control_attrition	attrition_dif
Started Baseline	9715				
Finished Baseline	5077	0.48	0.48	0.48	0.00
Started Endline	2061	0.59	0.59	0.59	0.00
Finished Endline	1894	0.08	0.10	0.07	0.03
Started Followup	569	0.70	0.70	0.69	0.01
Finished Followup	555	0.02	0.04	0.01	0.03

Given the high attrition and slow recruitment, the decision was made to focus on a pooled analysis of both countries. The treatment group experienced slightly higher attrition during surveys, possibly due to being asked more questions. However, the magnitude is relatively small.

Takeup

Table 3: Treatment Takeup

Dataset	Treated	Downloaded	Used	>1 Day	>3 Days	Used (%)	>1 Day (%)	>3 Days (%)
Serbia	678	379	189	85	18	27.9%	12.5%	2.7%
Bulgaria	342	182	116	46	11	33.9%	13.5%	3.2%
Pooled	1020	561	305	131	29	29.9%	12.8%	2.8%

The theory of change for Bebbo requires that individuals both download the app and use the app regularly. The evaluation assumption for this study was that by asking respondents to download Bebbo, a significant portion would download and regularly use the app.

29.9% downloaded and used the app once. This takeup is "significant" in the sense that we are powered in our study to measure an effect on this group. Unfortunately, most of those users never open the app beyond the first day. Only 9% of them open the app more than 3 days in the 4 weeks.

Table 4: App Engagement (30 days)

Dataset	Downloaded	Used (%)	>1 Day (%)	>3 Days (%)
Serbia	379	49.9%	22.4%	4.7%
Bulgaria	182	63.7%	25.3%	6%
Pooled	561	54.4%	23.4%	5.2%
Non-Study	40111	26.7%	13.6%	5.1%

We can compare the app usage of our participants to the app usage data from non-study participants to assess whether or not this is representative of real-world app usage. While we find that our study participants are more likely to finish onboarding (use) and more likely to use beyond the first day, they are equally likely to become regular users.

Power

Ex ante power analysis shows that we are well powered (above 80% power at a 1.25% significance level) with around 2000 participants in the pooled sample to find a medium effect size (0.5 standard deviations) at endline on those who downloaded and used Bebbo at least once (29.9% takeup).

In order to be equally powered at a takeup level of 2.8%, those that used the app more than three days in the treatment period, we would need 227,576 study participants. Even to measure the impact on those who used it at least two days would require 10,892 participants to be equally powered.

Assumptions

A core evaluation assumption failed: that asking participants to download and use Bebbo would cause a significant portion of them to use it regularly.

Thus, we will be able to evaluate the impact of downloading and using Bebbo once, but not the impact of "regular" Bebbo usage.

Findings

Findings

The analysis conducted on the effects of promoting the Bebbo app on caregivers with children aged 0-6 did not reveal any statistically significant impact on key outcomes of interest.

Findings

One of the main contributing factors that aligns with the theory of change is the low engagement levels of users with the app: without regular usage, we would expect no impact. Additional findings that might have contributed to the lack of measured significant impact includ:

1. There seemed to be a priming effect of the baseline survey, especially for questions related to knowledge, such as "when is your child's next vaccination due." This could imply that a prompt itself can change parent's knowledge and raises questions on how best to position Bebbo in regards to this outcome.
2. The majority of respondents scored well on the baseline assessment. If the respondents were representative of the general population, this would imply that most caregivers in these countries are already knowledgeable and following many good practices.

Findings: App Usage

Given that so few caregivers used the app, it seems important to ask the question: "who are the respondents who end up as app users?" We do so by regressing respondents' app usage activity against their characteristics at baseline.

Unfortunately, we are not able to accurately predict engagement from the measured baseline characteristics.

One notable predictor is "Activities Past 24h." One possible interpretation could be that there is a set of people who are not likely to spend time with their children but are likely to download and use apps. Unfortunately, the raw data is not reflective of a positive impact of this app on those people spending more time with their children. There is some additional evidence that those with young children are more likely to become regular users of the app as well as those who are parents (compared to other caregivers).
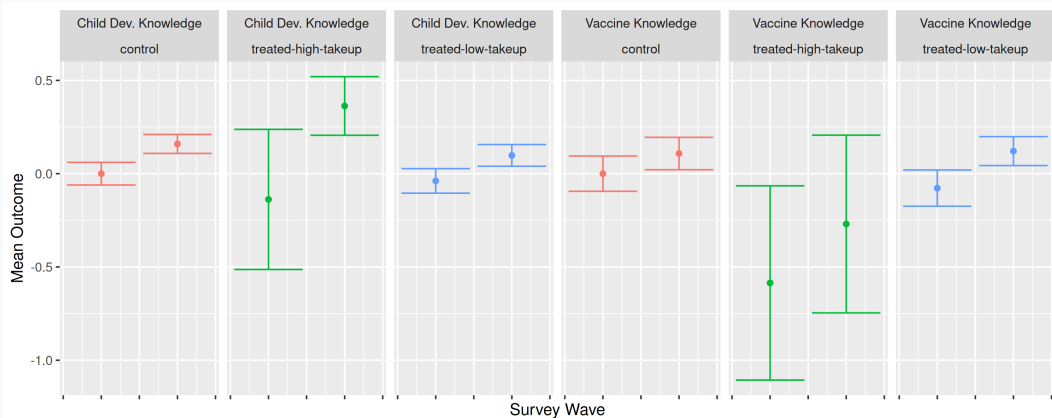
Suggestive Evidence

Suggestive Evidence

We are not able to detect a significant impact of treatment in this randomized control trial. However, it can be useful to examine the raw before-and-after data for suggestive evidence to be explored further in future research.

Note that before-and-after data does not represent the impact of usage, as users "self-select" into the group who uses the app. In other words: those who use the app more than 3 times do so for some reason, potentially because they are eager to learn or change. This group may have learned or changed with or without the app. Thus, what we are uncovering is correlation, not causation.

Suggestive Evidence: Knowledge and Awareness

There is evidence that awareness itself drives vaccine knowledge improvement. There is some suggestive evidence that regular app usage is correlated with increased knowledge and confidence in knowledge about child development stages.
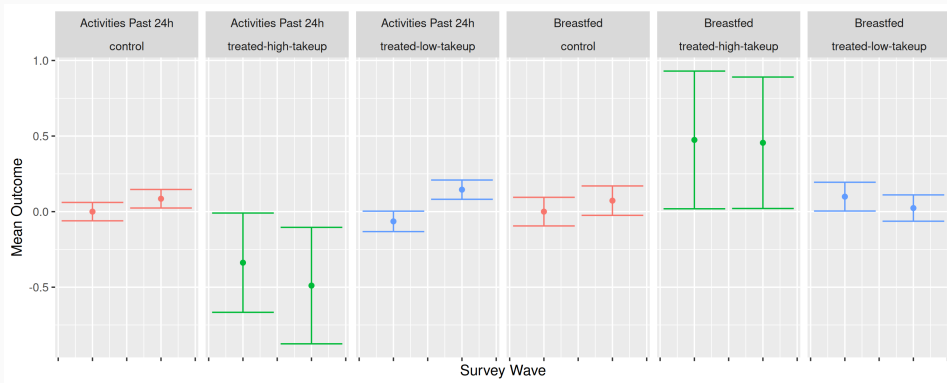

29

Suggestive Evidence: Confidence and Attitudes

There is evidence that regular app usage is correlated with improved confidence and attitudes towards physical punishment.
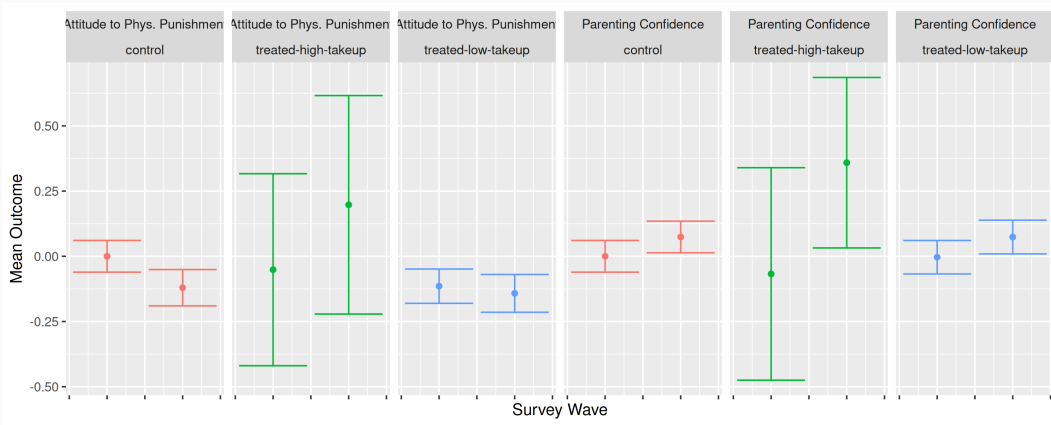
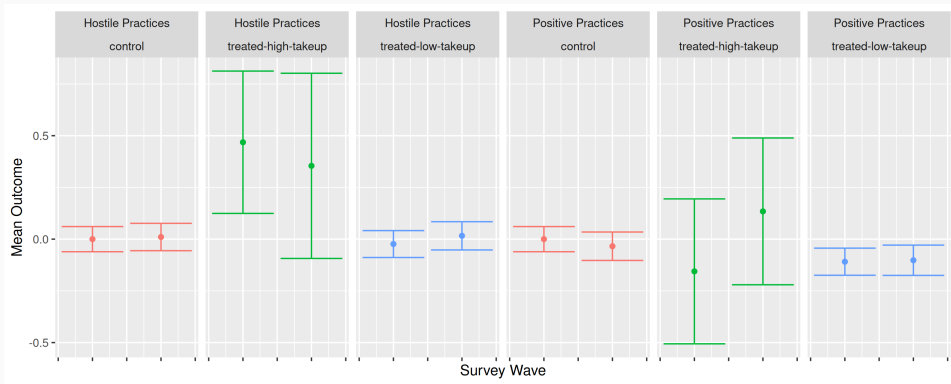There is evidence that regular app usage is correlated with spending less time with their child in important activities and seems independent of breastfeeding.

There is evidence that regular app usage is correlated with improved positive parenting practices.

Recommendations

Recommendations

For an app to influence behaviors and attitudes, it must be attractive and engaging enough to retain a large portion of the target audience over time.

Data indicate that a significant proportion of users tend to abandon the app after a short period. Specifically, 76% of users in our study who complied with the suggestion to download Bebbo abandoned it after the first day. This is in line with app usage data outside of our study, which shows that over 80% of users abandon it after one day.

We additionally find that only 3% of respondents used Bebbo more than 3 times in a 30 day period. Bebbo's theory of change is anchored in regular app usage by parents – the current usage rate is unlikely to lead to impact change.

Improve app targeting

The 5% of downloaders that remain as regular users, as well as other suggestive evidence, indicate that Bebbo may work for specific sub-groups. These groups need to be better understood to improve app outreach and uptake strategies. Unfortunately, we were not able to learn much about this subgroup in this analysis. It might be that new implementation modalities would be required to reach these groups. The current shift towards implementing through services providers (i.e., a hybrid strategy) might provide a solution, but it needs to be tested before being scaled up.

Adjust app focus based on context

Considering the high baseline scores of participants, it might be valuable (in places with already better than average parenting knowledge) to consider a focus on particular practices (i.e. breastfeeding) that are still lagging, or groups which are lagging in all practices.

Recommendations

Adjust app focus based on the comparative advantage to generate knowledge

The findings suggest that for many of the targeted outcomes, awareness itself is an effective intervention. This suggests that some outcomes (such as vaccine knowledge) might be more effective through other media and an app should focus on the outcomes that require sustained engagement.

Recommendations

Improve app usage

The current retention and usage rates are not sufficient for Bebbo to have any meaningful impact. Steps should be take to make the app more engaging for caregivers (and service providers considering the new implementation modality). These implementation issues should be addressed through formative and quantitative research (e.g., User experience (UX) studies to understand what would drive better user engagement, acceptability, accessibility, and reach) before further scaling the app.

Adopt a lean approach to evidence generation

Small, cheap, and continuous analyses should be used to better understand and improve the app's impact. Existing app usage data provides clear evidence of the biggest bottleneck: user onboarding and profile creation. Continuous user testing, lean UX techniques, and agile development methodologies can be used to address these issues quickly and affordably. Evidence generation can be used to continuously uncover and evolve the app, allowing it to adapt quickly and scale effectively.

Findings and Recommendations

In conclusion, while the promotion of the Bebbo app did not result in significant effects on the target population, the study provides valuable insights into the challenges of promoting and sustaining engagement with mobile apps for parenting.

Future interventions should consider strategies to enhance user engagement and retention to maximize the potential impact of mobile apps in promoting positive parenting practices and knowledge.