Impact evaluation of UNICEF's parenting app Bebbo: Results from a multi-country digital randomized controlled encouragement design

Nandan Rao

Advisor: Caterina Calsamiglia
Tutor: Francesc Trillas
June 17, 2024

Introduction

Background

- Early childhood development is critical for economic development, but major gaps exist worldwide (Black et al. 2018, Attanasio et al. 2022).
- Three main areas of focus: behaviors, attitudes, and knowledge of parents regarding caregiving and stimulation (Britto et al. 2017)
- Technology-assisted interventions have a history with mixed results, but still minimal evidence. Digital-only interventions may have lower engagement and worse retention than blended solutions, but few examples (Hall and Bierman 2015).

Need

- Understand how to create interventions that work at scale, improving early childhood development through impacting parents.
- Understand the effectiveness of mobile apps, as a digital-only solutions that improves parenting knowledge, attitudes, and behaviors.

Contribution

- Measure the impact of one particular early-childhood parenting app, as a part of understanding the effectiveness of this class of intervention.
- Deepen our understanding of the engagement and retention problem of digital-only solutions for population-level impacts.

TLDR;

- We measure the impact of a particular parenting app, Bebbo, with a prepost randomized encouragement design, on knowledge, beliefs, confidence, attitudes, and behaviors.

- We find no impact of this particular app. Primarily due to lack of engagement.

- We also find high ceiling effects and possible priming effects, which question the appropriateness of a general-purpose parenting app as an interventin.

- We suggest that more work is needed to effectively develop and measure digital interventions as public policy.

UNICEF Europe and Central Asia Regional Office (ECARO) developed the mobile parenting app, 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
2. Supporting parents for responsive caregiving and early intervention.

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.

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"

Measurement

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?"

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.

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

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.
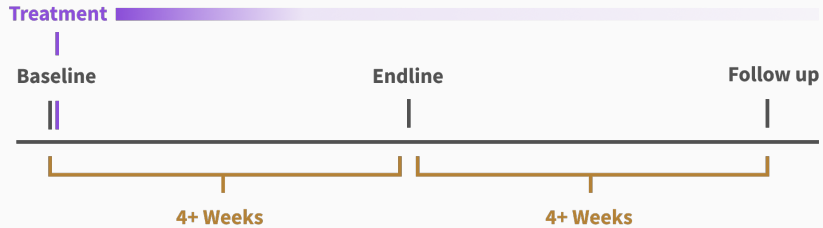
Figure 2: Study Design

Descriptives

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.
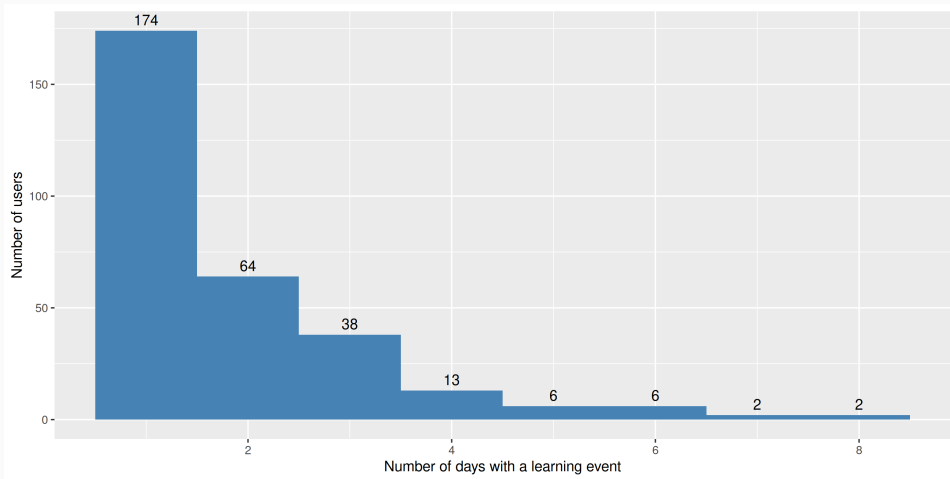
Figure 3: Days with Learning Event (Baseline - Endline)

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.

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.

Predicting Takeup

Table 4: App Usage (All)

	Used the App	Used More Than 1 Day	Used More Than 3 Days
	Dependent variable:		
	(1)	(2)	(3)
Is Woman	0.05(0.04)	0.01(0.03)	-0.02(0.01)
University Educated	0.07** (0.03)	0.04(0.02)	0.003(0.01)
Speaks Dominant Lang.	0.06(0.07)	-0.01(0.05)	0.01(0.03)
Is Parent	0.23*** (0.06)	0.13*** (0.04)	0.04* (0.02)
Child Age	-0.04(0.03)	-0.05** (0.02)	-0.02* (0.01)
Num. Children	0.01(0.06)	-0.01(0.04)	-0.01(0.02)
Parent Age	0.05(0.04)	0.03(0.03)	0.003(0.01)
Urban Area	-0.04(0.03)	-0.003(0.02)	0.01(0.01)
Child Dev. Knowledge	0.01(0.07)	-0.04(0.05)	0.03(0.03)
Parenting Confidence	-0.05** (0.02)	-0.04** (0.02)	0.01(0.01)
Attitude to Phys. Punishment	0.03(0.02)	0.004(0.01)	0.01(0.01)
Activities Past 24h	-0.05*** (0.01)	-0.04*** (0.01)	-0.02*** (0.01)
Positive Practices	0.04* (0.02)	-0.001(0.01)	0.005(0.01)
Hostile Practices	-0.01(0.02)	0.02(0.02)	0.004(0.01)
Constant	0.20(0.14)	0.29*** (0.10)	-0.02(0.05)
Observations	924	924	924
R^2	0.06	0.05	0.03
Adjusted R^2	0.05	0.03	0.02
Residual Std. Error (df = 909)	0.44	0.33	0.17
F Statistic (df = 14; 909)	4.21***	3.36***	2.14***

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

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.

Takeup: External Validity

Table 5: 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	42281	26.8%	13.8%	5.2%

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-post 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).

Asking participants to download and use Bebbo did not cause a significant portion of them to use it regularly.

Thus, we will be able to evaluate the impact of offering bebbo when it leads to downloading and using Bebbo once, but not the impact of "regular" Bebbo usage. This is an important consideration for any intervention with low takeup, which can often happen to digital-only interventions.

Findings

Model

We run the following regression model to measure the intent-to-treat effect (ITT) of assignment to the treatment arm:

$$y_i - y_i^b = \gamma_1 + \beta T_i + \gamma_2 X_i + \epsilon_i$$

Where y_i represents the outcome of interest for individual i measured after treatment, T_i represents the random treatment assignment, X_i a set of control variables and y_i^b represents the outcome of interest measured before treatment. The parameter of interest will be the treatment effect, β.

Note that due to the relatively large number of seperate outcomes (8), we adjust p-values of the treatment variable to control the false discovery rate (FDR), using Benjamini-Hochberg, reported as the "Adjusted Treatment p-value."

These two constructs, Vaccine Knowledge and Child Development Knowledge, both suffered from ceiling effects in the baseline survey (72% and 73% respectively). On top of those ceiling effects, they both potentially suffered from priming effects, as evidenced by the consistent improvement in the endline survey for all groups.

Results: Knowledge and Awareness

**Table 6:** Pooled: OLS - Endline - Knowledge and Awareness

	Dependent variable:	
	Vaccine Knowledge	Child Dev. Knowledge
	(1)	(2)
Treatment	0.05	−0.003
	(0.03)	(0.01)
Adjusted Treatment p-value	0.337	0.797
Observations	719	1,996
R^2	0.01	0.01

Note: *p$<$0.1; **p$<$0.05; ***p$<$0.01

Suggestive Evidence: Knowledge and Awareness

There is evidence that awareness itself (or time) 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.
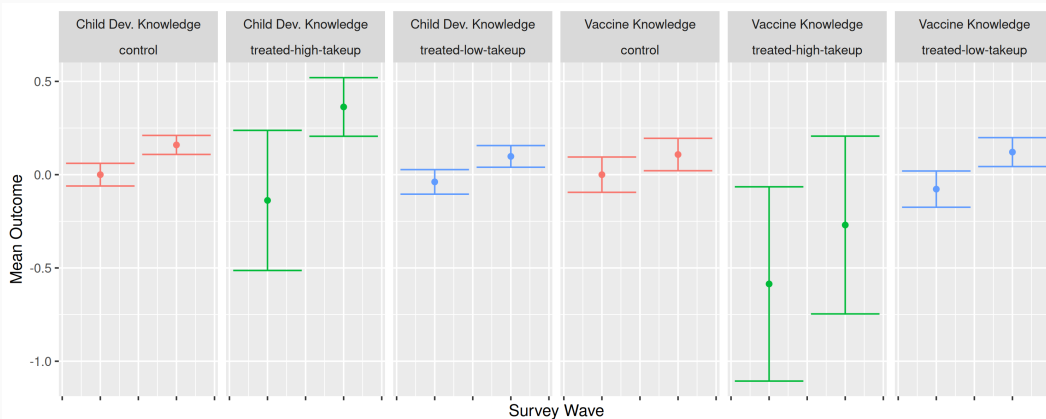
Results: Confidence and Attitudes

- Attitude Towards Physical punishment is a single question which asks if the parent believes the child needs to be physically punished.

- Parenting Confidence shows no significant impact in the regression analysis. The raw data shows suggestive evidence that those with lower confidence might be more likely to take up the treatment.

- The lack of a positive coefficient in the regression, however, might indicate that those in the control group were equally likely to take up either the control website or seek out information on their own in order to improve by endline.

Results: Confidence and Attitudes

	Dependent variable:	
	Parenting Confidence	Attitude to Phys. Punishment
	(1)	(2)
Treatment	0.01	0.08
	(0.03)	(0.04)
Adjusted Treatment p-value	0.797	0.157
Observations	1,972	1,961
R^2	0.01	0.01

Note: *p<0.1; **p<0.05; ***p<0.01

Suggestive Evidence: Confidence and Attitudes

There is evidence that regular app usage is correlated with improved confidence, as is time. App usage may be correlated with improved attitudes towards physical punishment.
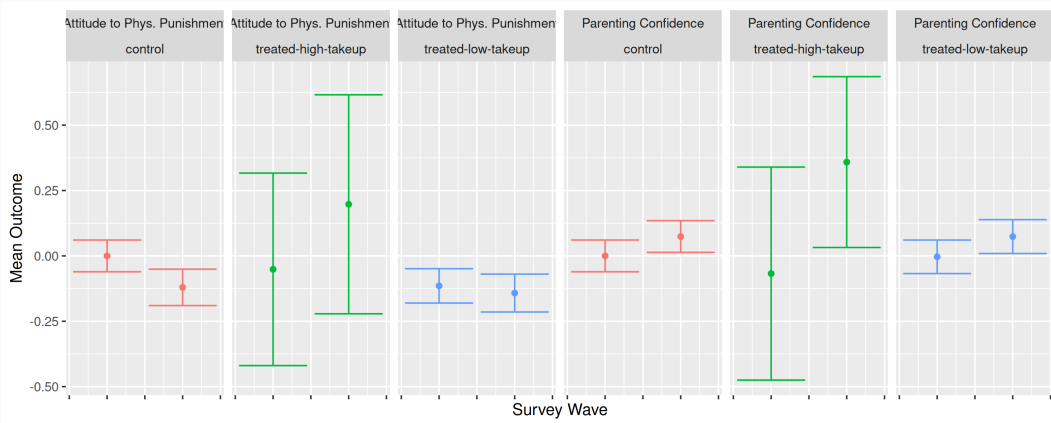
Results: Practices

- No significant effect was found for any of the behaviors.
- There is some evidence in the takeup regressions, and in the raw data shown here, that those who spend less time with their kids are more likely to use the app.
- The regression show that Activities Past 24h could have some suggestive evidence of impact, if not for the multiple testing correction. The raw data, however, shows that the much of the improvement is driven by those in the treated group who did not take-up the treatment

Results: Practices

Table 8: Pooled: OLS - Endline - Practices

	Dependent variable:			
	Breastfed	Activities Past 24h	Positive Practices	Hostile Practices
	(1)	(2)	(3)	(4)
Treatment	−0.02	0.12	0.03	0.03
	(0.03)	(0.05)	(0.03)	(0.03)
Adjusted Treatment p-value	0.714	0.157	0.618	0.618
Observations	682	1,903	1,904	1,900
R^2	0.02	0.01	0.005	0.01

Note: $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Suggestive Evidence: Practices

There is evidence that regular app usage is correlated with spending less time with their child in important activities and seems independent of breastfeeding.

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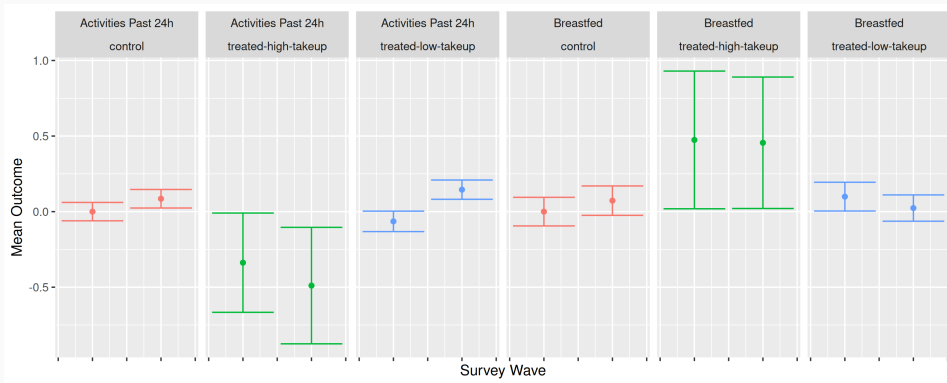
The analysis conducted on the effects of promoting the Bebbo app on caregivers with children aged 0-6 did not reveal any statistically significant impact on key outcomes of interest.

Findings: Explanation

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 include:

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.

Conclusions

Digital Intervention Needs

- For an intervention to have an impact on a population level, it is not sufficient to engage a small percentage of the population intensely

- This is in contrast to the majority of private-sector apps, who only need a small portion of the population to engage in order to make a successful business.

- Apps by their nature create upfront friction (download and install), which is rewarded with long-term, deeper engagement. Bad for population-level interventions?

- The priming effect we see implies that informational campaigns to promote awareness might have a significant impact—at significantly greater scale—than hiding that information inside an app.

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.