

Contents

1	Introduction	1
	About Bebbo	1
	What question does this evaluation answer?	1
2	Study Design	2
	Experiment Design	2
	Recruitment	2
3	Descriptives	3
	Respondent Characteristics by Country	3
	Construct Variables	3
	Baseline Balance	3

1 Introduction

About Bebbo

Parents everywhere are in need of information on various aspects of child development from reliable and validated sources as well as guidance on how to support the health and development of their children. However, services providing this sort of information and support are often non-existent or inaccessible for a lot of parents in many places. Often, service providers, even when accessible, might lack necessary knowledge and skills to respond to the questions and concerns parents might have.

Mobile apps are one of the most convenient and easy ways to access information about child development and parenting. However, parenting apps are mainly in English and provide a limited thematic content without a possibility for parents to familiarize with, track, and support all aspects of their child's health and development. In addition, these apps are, naturally, not adapted to contexts of individual countries. Many apps are not free of charge, which presents a significant barrier, particularly for the most vulnerable families. At the same time, the majority of the existing apps operate only in online mode requiring good internet connectivity that is lacking in remote and rural areas.

To support parents to receive timely and quality guidance even when direct contact with service providers is not possible and overcome barriers in access to localized digital solutions with verified content, UNICEF Europe and Central Asia Regional Office (ECARO) developed a mobile parenting app, Bebbo. The mobile application also supports the most vulnerable parents/caregivers with lower education level, in terms of the navigation modalities, off-line operability and selection of the core content. 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 app provides users information and interactive tools to help nurture and aid their child'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, stimulating, and learning.

What question does this evaluation answer?

The design of the study is set up to answer the following question in the positive: is asking parents to use Bebbo an effective policy to improve the parenting knowledge, attitudes, and practices of the general population in Bulgaria and Serbia? Note that the study cannot fully answer the question in the negative, it cannot prove that this intervention is ineffective, it can only fail to measure its effectiveness.

In order to understand this study, it's important to break down the question we are answering and understand how the design of the study limits or defines each component:

Asking parents to use Bebbo Study participants were randomly divided into two treatment groups. Each group was surveyed at baseline and then "treated" through an invitation to engage with parenting material. One group was asked to visit a website (this is the "treatment as usual", TAU, or the "control" group). The other group was asked to download Bebbo (the "treated" group). By comparing Bebbo to the existing TAU, we are asking the question: "does this new treatment offer something above and beyond the already existing treatments which parents might presumably already be asked to do?"

Why ask? This study follows a randomized encouragement design. This design is used whenever the following two conditions occur:

1. One is interested in the impact of a treatment on a population where individuals can choose whether or not to take the treatment (the “compliers”)
2. The compliers and non-compliers might have different reactions to the treatment.

This design is not useful if the reaction to the treatment is independent of willingness to comply. This design is also not useful if one wants to know the theoretical impact that the treatment could have on those who choose to not take the treatment.

Effective An effective policy, in this study design, is one that can be measured to: Improve the way in which a treated individual responds to a set of survey questions, the second time they are asked, compared to the first time they were asked, and then compared to the treatment as usual.

Does so when the treatment is applied to the individual after being asked the survey questions the first time and 4-6 weeks before answering the survey questions the second time.

General Population The general population in these countries had significant pre-exposure to Bebbi before the beginning of the study. This population is different from a population that has no exposure to Bebbi. One would expect that asking people to download and use Bebbi would have a bigger impact on a population that has no exposure to Bebbi.

Additionally, no care was given to single out any particular subset of the population that might benefit the most (or the least) from Bebbi, nor those who would be most likely to use Bebbi.

What do we find?

We do not find evidence that asking this population to use Bebbi has any impact beyond that of asking them to visit a parenting website.

Given the design of the study, we can say that the following facts may have contributed to the lack of evidence of impact and are interesting facts in their own right, when considering the potential impact of this intervention:

1. The population was already very “good” in regards to the outcomes of interest. We measured the improvement of parents over time, under treatment, but many could not improve from their baseline scores, which were perfect.
2. Participants improved from the first questionnaire to the second questionnaire, regardless of treatment arm and regardless of compliance. This seems to imply that the very act of asking the questions improves the way that parents answer them.
3. Very few people complied with treatment and used Bebbi. Of those who were asked to use the app, 28% used the app, 12% used the app more than one day, and only 3% used the app more than three days. With a small percentage using the app intensively, the impact on the entire population will be too small to measure with the size of this study. Significant pre-exposure could have led to the low initial compliance (the 28

2 Study Design

Experiment Design

This study follows a pretest design (Clifford, Sheagley, and Piston 2021) in which we measure the outcomes of interest before treatment (in a baseline survey) and after treatment (in an endline survey). We add an additional survey after the endline, referred to as a follow up, to look for longer-term impacts and test the impact of continued app usage.

Study participants are randomized, from the beginning, to one of two conditions:

1. **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 Bebbi and use it regularly, being encouraged that doing so will help them with their parenting.
2. **Control.** Participants in the control condition were told that there was one more step to qualify for the study and were then asked to visit a parenting website and use it regularly, being encouraged that doing so will help them with their parenting.

App usage in the treatment group was tracked, however website visits in the control group were not tracked.

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. In exchange for participating in the study, they were told they could receive gift cards worth up to 12 USD (in their local currency). See [REF pictures] for examples of the ad material used for recruiting.

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 and sent the endline survey 4 weeks after completing the baseline survey.

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 with the Virtual Lab Facebook page, which did not contain any content or information related to this study. Consent was provided via chat, as well as all answers to the survey questions and the treatment condition. Gift cards were also provided via chat, using the Tremendous gift card platform to provide Visa international prepaid cards.

3 Descriptives

Respondent Characteristics by Country

Table 1 provides the baseline characteristics of the respondent population, separated by country.

Generally speaking, most respondents were themselves parents (not grandparents or other caregivers), women, under 35 years of age, and spoke the dominant language of the country at home. A little over half had children 0-2, compared to 2-6 years of age.

Respondents in Bulgaria were more likely to have a university education (42%) compared to those in Serbia (29%).

Table 1: Baseline Respondent Characteristics

Variable	Value	Bulgaria	Bulgaria %	Serbia	Serbia %
woman	1	1418	0.83	2102	0.80
university	1	725	0.42	748	0.29
dominant_language	1	1571	0.92	2488	0.95
is_parent	1	1485	0.87	2374	0.91
age_flag	2-6	935	0.55	1561	0.60
children_count	4+	70	0.04	279	0.11
parent_age_flag	Over 35	365	0.21	550	0.21
urban	1	1059	0.62	941	0.36

Construct Variables

The...

See table 2

Baseline Balance

To test for balance between our randomly assigned treatment and control groups, we run an omnibus test, following Hansen and Bowers (2008), to observe standardized differences at baseline and the associated omnibus p-value. Results are found in table 4.

The most notable variable is the practices_24 and was_breastfed, which differ a bit between the groups although they are still under 0.1 standard deviations in their difference. The omnibus test, with a p-value above 0.13, implies that the two groups are not different in a statistically significant manner.

Table 2: Construct Variable Mapping

Domain	construct_variable	question
Knowledge and awareness	health_knw	I know which vaccine {{field:child_name}} needs to take next.
Knowledge and awareness	dev_knw_recog	I would be able to recognize if {{field:child_name}} lags behind in social-emotional development (expressing and recognizing feelings and emotions, engaging in interactions, etc.).
Knowledge and awareness	dev_knw_recog	I would be able to recognize if {{field:child_name}} lags behind in cognitive development (mental development, intellectual development).
Knowledge and awareness	dev_knw_recog	I would be able to recognize if {{field:child_name}} lags behind in physical development.
Knowledge and awareness	dev_knw_recog	I would be able to recognize if {{field:child_name}} lags behind in language development.
Confidence and attitudes	confidence	How confident do you feel in your ability to deal with {{field:child_name}}'s emotions?
Confidence and attitudes	confidence	How confident do you feel in your ability to respond properly when {{field:child_name}} misbehaves?
Confidence and attitudes	attitude	Do you agree that in order to bring up, raise, or educate a child properly, the child needs to be physically punished?
Confidence and attitudes	caregiver_well_being	How often can you handle stressful parenting situations successfully?
Practices	was_breastfed	Has {{field:child_name}} been breastfed in the last 24 hours?
Practices	practices_24	In the past 24 hours, did you read books or look at picture books with {{field:child_name}}?
Practices	practices_24	In the past 24 hours, did you tell stories with {{field:child_name}}?
Practices	practices_24	In the past 24 hours, did you sing songs (including lullabies) to or with {{field:child_name}}?
Practices	practices_24	In the past 24 hours, did you take {{field:child_name}} outside the home?
Practices	practices_24	In the past 24 hours, did you play with {{field:child_name}}?
Practices	practices_24	In the past 24 hours, did you name, count or draw things with or for {{field:child_name}}?
Practices	practices_agree	When {{field:child_name}} and I play together, we laugh a lot.
Practices	practices_agree	I joke around with {{field:child_name}}.
Practices	practices_agree	I often smile when I'm around {{field:child_name}}.
Practices	practices_agree	{{field:child_name}} and I play together on the floor.
Practices	practices_hostility	I snap at {{field:child_name}} when he/she gets on my nerves.
Practices	practices_hostility	When {{field:child_name}} upsets me, I lose my patience and punish him/her more severely than I really mean to.
Practices	practices_hostility	When {{field:child_name}} does something wrong, I sometimes threaten him/her.
Practices	practices_hostility	I sometimes make fun of {{field:child_name}}.

Table 3: Outcome Construct Descriptives Pooled Baseline

Subdomain	name	mean	median	min	max	sd	prop_na
Attitude	attitude	3.13	3.0	1	4	0.84	0.00
Child development knowledge	dev_knw_recog	0.86	1.0	0	1	0.28	0.00
Confidence in parenting	confidence	3.34	3.5	1	4	0.65	0.00
ECD engagement	practices_24	4.92	5.0	0	6	1.23	0.00
Health knowledge	health_knw	0.72	1.0	0	1	0.45	0.58
Nutrition	was_breastfed	0.37	0.0	0	1	0.48	0.58
Responsive parenting Hostility	practices_hostility	3.04	3.0	1	4	0.69	0.00
Responsive parenting PA	practices_agree	3.20	3.5	1	4	0.75	0.00

Table 4: Baseline Balance Pooled

	control_mean	treatment_mean	standardized_diff	z_score
health_knw	0.73	0.71	-0.06	-1.20
dev_knw_recog	0.86	0.85	-0.04	-1.29
confidence	3.33	3.34	0.01	0.37
attitude	3.14	3.11	-0.04	-1.23
was_breastfed	0.35	0.39	0.08	1.79
practices_24	4.97	4.87	-0.07	-2.46
practices_agree	3.20	3.19	-0.01	-0.39
practices_hostility	3.03	3.04	0.004	0.12
(health_knw)	0.42	0.43	0.02	0.67
(was_breastfed)	0.42	0.43	0.02	0.61

Overall P-Value: 0.134132865981628