BebeCODE: Collaborative Child Development Tracking System

Seokwoo Song **Juho Kim** **Bumsoo Kang** **Wonjeong Park**[†] **John Kim**
School of Computing, KAIST, Daejeon, Republic of Korea
[†] Department of Communication Disorders, Ewha Woman University, Seoul, Republic of Korea
{sukwoo24, juhokim, steve.kang, jjk12}@kaist.ac.kr [†]thisisforwj@ewha.ac.kr

ABSTRACT
Continuous tracking young children's development is important for parents because early detection of developmental delay can lead to better treatment through early intervention. Screening tests, often based on questions answered by a parent, are used to assess children's development, but responses from only one parent can be subjective and even inaccurate due to limited memory and observations. In this work, we propose a collaborative child development tracking system, where screening test responses are collected through collaboration between parents or caregivers. We implement BebeCODE, a mobile system that encourages parents to independently answer all developmental questions for a given age and resolve disagreements through chatting, image/video sharing, or asking a third person. A 4-week deployment study of BebeCODE with 12 families found that parents had approximately 22% disagreements about questions regarding their children's developmental and BebeCODE helped them reach a consensus. Parents also reported that their awareness of their child's development, increased with BebeCODE.

ACM Classification Keywords
H.5.3 Group and Organization Interfaces: Computer-supported cooperative work; J.3 Computer Applications: Life and Medical Sciences

Author Keywords
Children; families; early detection; developmental tracking; parents' collaboration; mobile system

INTRODUCTION
Screening tests based on parents' reports are commonly used to identify possible developmental delays of children. These tests involve asking parents a series of questions about their children's development including language use, physical development, cognition, and so forth [20]. To answer developmental questions, parents need to recollect specific activities of their children (e.g.,"child can differentiate between big and small objects" or "child says hello to a familiar adult" [43]). Properly identifying possible developmental delays is critical because early diagnosis and intervention can result in better treatment of delays [6].

Manually tracking developmental milestones over an extended period of time can be a difficult task for parents. In addition, parents can fail to recollect new developmental achievements of their children or even not notice particular developments [24]. To overcome these limitations, prior studies [25, 39] have proposed systems that can help parents keep track of their child's developmental milestones over an extended period of time. However, these studies assumed that a single parent is responsible for tracking development.

Although screening tests answered by a single parent are commonly used in assessing children's development [20], some inaccuracies due to parents' limited observations or their subjectivity and misinterpretation of questions are known issues [35]. For example, a parent might exaggerate their child's development or underestimate it due to poor recollection or limited observation even if the child has achieved a particular development milestone. Another challenge can be vague questions asked in the developmental questionnaire. For example, considering a question asking whether a 'child can have a "simple"' conversation [43], one parent might think that their child can it well, while the other parent might think that their child cannot do it. In addition, other adults or caregivers might have a better awareness of the child's development. For example, a daycare teacher may have a better understanding of a child's social skills and their ability to interact with other children. Grandparents or babysitters might also help observe a child's developmental activities. As a result, a third person might be in a good position to provide accurate answers to development-related questions [33]. Prior studies [10, 34] in the field of developmental psychology have found that having multiple informants lead to more reliable results in assessing a child's development in comparison to having a single informant.

In this work, we propose a *collaborative* approach to monitoring children's development through BebeCODE – a mobile system for **CO**llaborative Child **DE**velopment tracking. BebeCODE encourages parents to assess children's development. The particular characteristics of BebeCODE include the following: (1) both parents are required to answer all questions in a developmental area for a given age and need to reach *consensus* upon disagreement to view the test results; (2) to resolve disagreements, parents can chat, share images/videos or consult a third person to get their opinion, and (3) BebeCODE sends push notifications to remind parents to re-evaluate their child's developmental status after a time-out period as the child grows, and sends progress status notifications for successful collaboration between parents.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

DOI: https://doi.org/10.1145/3173574.3174114

To evaluate BebeCODE, we conducted a 4-week deployment study with 12 families with children ranging in age from 10 to 50 months old. The study found that when both parents actively participated in answering questions independently, they had approximately 22% disagreements in their answers. BebeCODE helped the parents reach consensus via three types of features on the questions about which there was a disagreement, and it enabled them to continuously track their children's development via notifications during the study. Seventy-six percent of the total disagreements were resolved, with chatting and simple conversation being the most frequent channel for resolution; 8% were resolved by sharing of images/videos, and 10% by asking a third person for their opinion. The parents also reported that use of BebeCODE increased their awareness, understanding, and interest in child development. Furthermore, some reported that the system led them to have more conversations between parents about the status of their child's development. The results of the study also suggested areas that could be improved for the design of collaborative assessment applications.

RELATED WORK

This work builds upon previous research on child development assessment, its technology support, and collaboration systems for health.

Early Diagnosis/Intervention of Developmental Delays

Developmental delays significantly affect various aspects of children's lives, including their subsequent academic performance [4]. Prior research found that early intervention had a positive impact on improving children's interaction skills with their parents [14], social-emotional skills [2], and their academic performance after school age [15] for autism spectrum disorders, cerebral palsy, and other developmental disorders [11].

In a clinical setting, experts directly assess children's development with various testing tools such as the Bayley Scales of Infant and Toddler Development [1] or the Denver developmental screening test [13]. However, children's actual development age can be underestimated because of children's unfamiliarity with the clinical setting. To overcome the limitations of direct assessment, indirect assessment based on parents' reports are commonly used together with direct assessment [13, 20]. While assessment based on parents' report can be as reliable as a direct assessment by experts, some parents' reports can be subjective and even inaccurate, limited by their interpretation or observation [35, 18]. To get more reliable assessments from parents, prior studies in the field of developmental psychology [10, 34] found that observations from multiple informants led to more reliable results in assessing children's development in comparison to a single informant. In this work, we pursued a similar approach through collaborative assessment of children's development

Technologies for Assessing Children Development

Child development tests commonly evaluate children's motor, recognition, and language skills as well as their socialiability [20]. Many researchers have proposed systems to assess child development for a particular domain. Measuring grasping actions and force is an important motor skills for children, and a sensor-embedded toy was proposed to analyze children's grasping action [7, 42]. TalkLIME assesses children's language development by comparing the number of average utterances of a child to that of others in the same age group, and provides real-time feedback to improve the parent-child interaction and reinforce the parent-training intervention [38, 19]. Monitoring the kicking patterns of infants can be used to track general development. To estimate the pose of an infant's leg, a computer vision-based approach has been used [36]. To detect autism spectrum disorder, robots are often employed to evaluate children's sociability [3]. These works have focused on automating assessment in a single domain of child development, and the challenge in assessment is recognizing children's activity.

Screening tests based on parents' reports are commonly used to assess children's overall development. Kientz et al. proposed Baby Steps for tracking the developmental progress of young children by parents [24, 25]. BabySteps for Twitter was proposed to actively involve social media in the tracking of children's development, in which parents respond to tweets about developmental milestones [39]. BabyStep Text [40] supports both parents' participation in tracking children's developmental, but it does not consider the possibility of differences in opinion between parents. Our approach is also focused on assessing children's development based on parents' reports. However, we propose a mobile system to encourage collaboration between parents or caregivers to overcome the limitations of only one parent participating in the screening tests.

Family-Centered Collaborative System

Prior research examined how to support health-related collaboration between caregivers and patients for better care by an improved understanding of each other [8, 27, 32, 29]. Recently, the HCI community has proposed systems for family-centered collaboration between parents and children. Toscos et al. proposed health monitoring technology to share a child's glucose levels with their parents [41]. Hong et al. designed a system to support teens in partnering with their parents for life-long treatments of a chronic disease [16]. Since the health of parents and children is interconnected in the family, a family-centered health monitoring system was proposed [31].

Although much of the related work has focused on sharing users' personal information to close the gap in their understanding through collaboration between caregivers and care recipients, our design focuses on resolving differences of opinion in a collaborative system between parents when assessing children's development to reach a consensus.

PRELIMINARY STUDY

We conducted a preliminary study to find opportunities for collaboration between caregivers for better assessment of child development. Mothers and fathers tend to have different roles in raising their children, and this results in different levels of understanding of the children [37]. When parents spend time with preschool children, mothers tend to teach and engage in empathic conversations, while fathers tend to behave like

Table 1. List of interviewees who participated in the preliminary study

	ID	Descriptions	Experience (child's age)
Group 1	S1	Child developmental specialist & Child psychologist	15y
	S2	Child developmental specialist	2y
	S3	Child developmental specialist & Play therapist	3y
Group 2	T1	Daycare teacher	6y
	T2	Kindergarten teacher	10y
Group 3	P1	Father	18m
	P2	Father	41m
	P3	Mother	37m

children by engaging in physical play [21]. In addition, daycare teachers closely observe children's social interactions that are hard to observe at home due to limited opportunities for peer interaction. If the caregivers of a child can share their unique observations, it could enable more accurate tracking of the children's development than reports based on a single caregiver's response [33].

Interviews with Developmental Specialist and Caregivers

We interviewed child developmental specialists, daycare teachers and parents of preschool children in a semi-structured interview to explore the possibility of caregivers' collaboration in assessing children's development (Details of the participants are summarized in Table 1). Each interview lasted approximately one hour. Topics covered by the interviews include the following: (1) understanding child development assessments conducted in public health clinics, (2) communication between daycare teachers and parents, (3) differences between a mother and father in the understanding of their child's development. The key topic areas of the questions we asked each group are summarized as follows:

- Child developmental specialists (Group 1)
  1. General methods for checking children's development.
  2. Advantages of parent-based child development screening tests when father and mother participate together
- Daycare center teachers (Group 2)
  1. Overall method of child development management in their daycare center
  2. Frequency of communication with children's parents
  3. Burden of personal contact with children's parents
- Parents of preschool children (Group 3)
  1. Parents' differences in style of childcare
  2. Experience of differing perception of a child's development between the father and the mother

All interviews were recorded and transcribed. We used open coding and iterative clustering to investigate emerging themes. We then identified core design factors to enable collaborative assessment of children's development from parents and/or other caregivers.

Interview Results

Parent involvement in clinical assessment

In the clinical assessment of a toddler, the participation of both parents is preferred in assessment as well as in treatment [9]. The child developmental specialists we met all preferred parents to come together when consulting child's development. When parents talk together, it helps to reduce the uncertainty of one parent: *"It is common that some parents subjectively think that their child can or cannot do something, an opinion with which I cannot agree, and some parents misunderstand the developmental questions (S2). (...) I encourage parents to visit together because parents' opinions often differ. (...) believe it is more accurate when parents talk together about the child's developmental activity that happened a while back (e.g., when did your child start to walk?) (S1)* Even though the experts emphasized the importance of both parents' involvement, because of the parents' busy schedule, it was rare for both parents to visit together: *"I recommend both parents to visit together but the mother usually brings the child alone because the father has to work. There are some cases in which a grandparent brings the child when both parents are working (S3)."*

Differences in children observations among parents

According to OECD's (Organisation for Economic Co-operation and Development) 2015 report, childcare participation differed between mothers and fathers in most of the OECD countries [30, 22]. In case of a single income family, one parent works while the other parent is usually dedicated to childcare: "Because my spouse works, I do most of the childcare (...) I periodically take a child development test and visit the hospital alone with my child (P3)." Primary caregivers spend most of the time with their child, so they know more than their spouse. However, different style of participation in childcare can result in different observation of a child, which results in different opinion in assessment of children development: *"My wife doesn't give a dangerous item to the child such a scissors or a pencil, but I do (P2)," "I usually play with my child with physical activity unlike my wife (...) Because I cannot participate in childcare during weekdays, I try to participate more over weekends (P1)"*.

Understanding differences of opinions between parents

Based on the interviews with experts and parents, we classified reasons why differences of opinion occur between parents on the child's development questions. The differences can be classified into one of the following three categories:

- **Single-parent observation:** The most obvious difference occurs when only one parent has observed their child's activity but their spouse has not. This difference in opinion can be easily resolved if the parent who has made an observation lets the spouse know or has a photo or video that captures the activity.
- **Different recollection:** Both parents might have seen their child's activity, but one parent might not remember it. This difference can be easily resolved if the other parent reminds the spouse (e.g., "he/she did it at home last weekend").
- **Different criteria for scoring or assessing**: There are some subjective questions on the children development test, such as "child can run naturally," but the definition of "naturally" can differ between parents. This difference in opinion can be resolved through discussion, and in some cases, a third person's comments might help resolve the difference.

Opportunity from daycare teacher's involvement

Parents and daycare teachers can actively communicate about children's daily activity through applications such as Kidsnote [23]. The daycare teachers who we interviewed also write a daily diary to share with parents about their child's status, recording several details, such as moods, meals, naps, condition, and bowel movements, and they write a weekly diary summarizing how each child spent a week. Daycare centers also provide a variety of play activities that naturally exhibit children's developmental progress. Because daycare centers are places where many children stay together, social activities of children that are difficult to see at home can lead to unique observations: (T1) *"I spend more than 8 hours per day with children. I would better know about children's development than their parents in some parts (...) In periodical interviews with parents, their most common question is about their children's sociability."*.

The daycare teachers agreed that they know better than parents about some of the characteristics of their children. However, the teachers felt uncomfortable having much personal contact with parents to talk about their children because it is an extension of their work after the workday: (T2) *"I never share my personal phone number, because the parents would frequently contact me asking about their children even late at night."* Also, although teachers know children's problems well, they were cautious about openly sharing this knowledge with parents because some parents blame teachers for the problem: (T2) *"It is not easy to be honest with parents when a child has a problem because if I tell them, the parents would frequently contact me and would even blame me for the problem."*

Design Implications

Based on the preliminary study, design implications were drawn to design a collaborative tracking system for a child development test. First, both parents have to participate in assessing child development to overcome the limitations of only one parent reporting. Second, the system should help parents resolve their disagreements in their answers to developmental questions. Third, a third-person's information is useful but the system should consider a less disruptive way to ask them about child development. Finally, the system for tracking child development should provide a long-term way to continuously track child development [24].

BEBECODE SYSTEM DESIGN

In this section, we describe the features and implementation of BebeCODE, which focuses on successful collaboration of parents in assessing their child's development. Our main design concept is parents' collaboration. They must both participate in the assessment to complement each other's lack of knowledge about certain points regarding their child's development. BebeCODE supports parents to resolve their disagreements to reach a consensus. In addition, whereas parents are required to participate, a third person's participation is optional to reduce their burden. The overall BebeCODE usage process is shown in Figure 1 and summarized below.

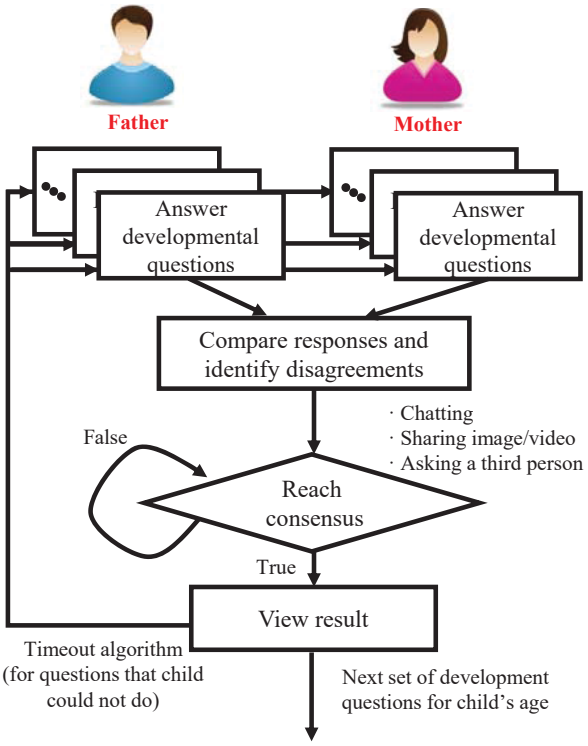

Figure 1. Overall usage process diagram. The timeout algorithm is only applied to developmental questions for which a child could not achieve and those development questions are repeated through the process.

(1) Both parents should answer all developmental questions independently (Parents' responses that their child cannot do a certain activity or task disappeared after N days to ensure validity.)
(2) The questions about which there is disagreement can be shown.
(3) The disagreements should be resolved via chatting, sharing of images/videos, or asking a third person for their opinion to reach a consensus.
(4) When there are no incomplete or questions for which there is disagreement, the assessment results are shown.

Questions for the Developmental Screening Test

We used the Korean Developmental Screening Test (K-DST) for Infants and Children [43] on BebeCODE, which was developed in Korea by the Pediatrics Society for screening of young children (4 months to 71 months) based on parents' reports to detect early signs of delays. Parents can screen their children's development with K-DST, including gross motor skills, fine motor skills, cognition, language, sociability, and responsibility. It requires parents to choose among the following in response to each question *"absolutely can't do"*, *"can't do well"*, *"can do"*, or *"can do well"*. The questions change every few months as the expected development of the child changes. We define *disagreement* or *different opinions* of parents on the questions when one parent chooses *"absolutely can't do"* or *"can't do well"* and the other choose *"can do"* or *"can do well"* because such conflicting responses cannot both be accurate.

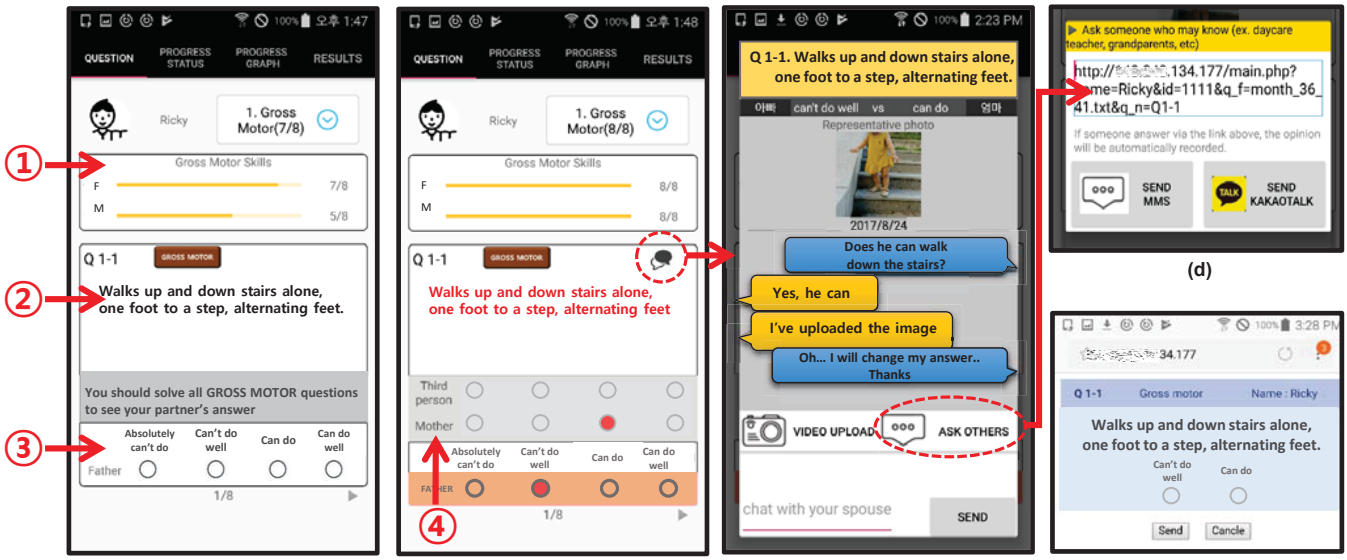

Figure 2. (a) Response input interface that includes ① response progress, ② developmental questions, and ③ response input, (b) disagreement of parents on this question and ④ other's opinion. (c) parents can communicate by chat and upload image/video to share child's activity for reaching consensus on this question (d) URL for asking a third person, (e) web-based developmental question and response for third person. The interface is translated from Korean.

Collaboration Between Caregivers

Figure 2 shows the overall interface to support the collaboration of parents and caregivers to assess a child's development. The procedure of using BebeCODE is summarized as follows.

(1) Answer all questions in a developmental area: Parents have to answer all question to see others' response for each developmental section such as gross motor, language, and cognition. Because the others' response can affect one's answer, we intentionally designed the interface so that parents would think about their children's development first before checking others' responses. By sliding the screen, the questions can be changed, and the progress bar on the upper side shows the number of questions answered (Figure 2(a)).

(2) Compare to others' answers: After parents answered all questions in a section, they could see each other's answer. If the opinions of the parents are different, the color of the question changes to red, and a discussion icon blinks to indicate that there is a disagreement between the their responses. In addition, parents can input a third person's answer on this screen if a third person's opinion is known, but the third person's opinion does not affect the disagreement question; the answer from a third person can be selected either *"can't do well"* or *"can do"* to minimize their weight in assessing the child. (Figure 2(b)).

(3) Discuss disagreement questions to reach consensus: When the discussion icon is selected (Figure 2(b)), the discussion interface is shown (Figure 2(c)). Parents can discuss their differences and attempt to resolve their differences by chatting or by sharing images/videos to persuade their spouse.

(4) Ask a third person (optional): When asking a third person about their child's development, users can send a URL to someone who may be able to answer the questions. On the webpage, the respondent can answer either *"can't do well"* or *"can do"*, and the result is automatically saved in the database server. This web-based approach does not require the third person to install BebeCODE (Figure 2(d)).

Notification for Periodic Re-evaluation

While conventional parent screening tests are often answered all at once, BebeCODE enables parents to answer questions incrementally through the mobile system. This also provides the opportunity to re-evaluate screening test questions for which the parents responded *"can't do."* Since child development changes continuously, responses from the previous week or month might not be valid, as the child's development could have changed from *"can't do"* to *"can do."* As a result, BebeCODE implements a *timeout* algorithm where the system pushes a new message to the parent to re-evaluate the child development questions for each task that the child was not able to achieve. Based on a discussion with child development specialists, we used a 7-day timeout algorithm to ensure that the re-evaluation messages were not too frequent. This approach enabled periodic re-evaluation of child development, which is very difficult to achieve with a conventional paper-based parent screening tests.

Functionality for Successful Collaboration

For successful collaboration between parents, knowing how their partner is doing is important [26]. BebeCODE provides a progress status bar and notifies parents by sending push messages when their spouses have made meaningful progress, such as when all questions in a section are answered, chat messages or images/videos are uploaded, or disagreement questions are resolved (Figure 3(a)). To check the final test results, parents have to answer all questions in a section, and there have to be no questions for which there is disagreement between the parents. The weekly developmental score change is shown (Figure 3(c)) to express a child's development, and

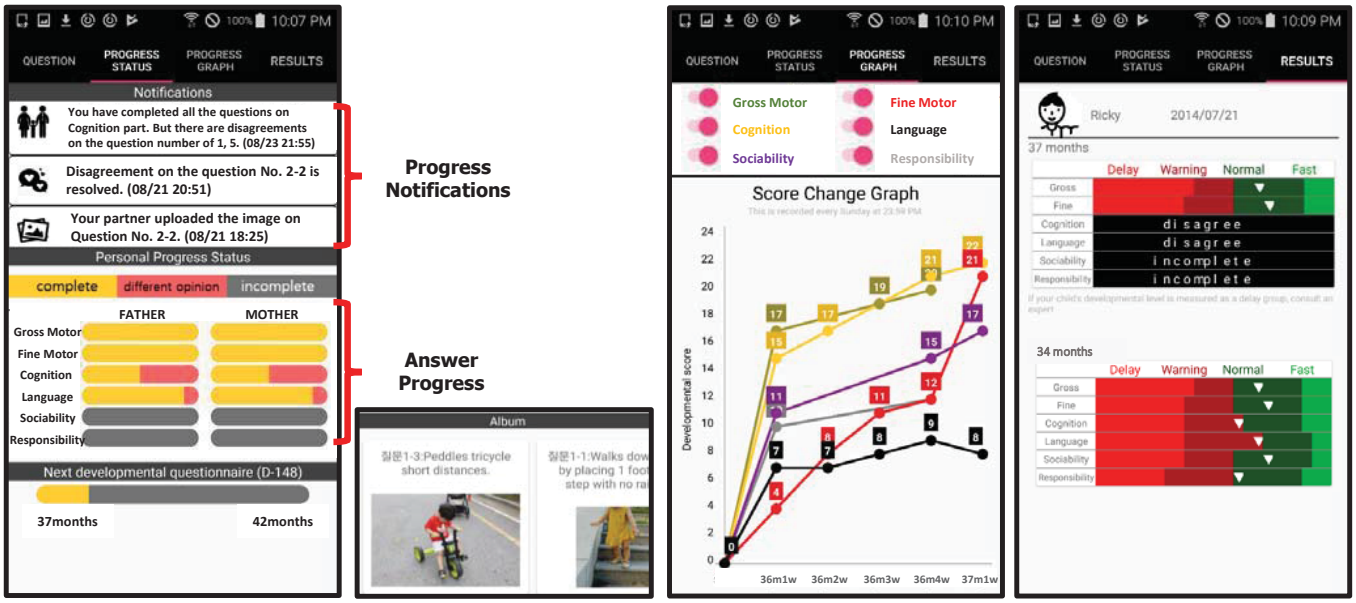

Figure 3. (a) Notification messages and answer progress, (b) uploaded images for each question, (c) the line graphs showing the child's weekly development progress per area, and (d) the summary of developmental test results. The interface is translated from Korean.

the screening test result is shown with detailed developmental level (Figure 3(d)).

BebeCODE Implementation

BebeCODE was implemented on the Android operating system. The server was developed in Java to save parents' development answers, uploaded images/videos, and application usage pattern records. It also automatically sends push notifications to participants when the proper progress is detected, and it invalidates recorded answers that are checked as *"can't do"* after 7 days to ensure response validity. In addition, a web server is run to create a page that is sent to a third person to get their opinion.

EVALUATION: FIELD DEPLOYMENT STUDY

To test the feasibility of BebeCODE for parents, we conducted a field deployment study with 12 families with young children. The study sought to determine whether BebeCODE can help parents resolve disagreements and track their children's development.

Participants

To recruit parents of young children between 4 months and 71 months old [1], we posted recruiting posters on childcare-related online communities. We asked applicants to answer a pre-study survey that asked about their demographic information and parenting environment. For each child, both parents had to agree to participate in the study. Since BebeCODE was implemented on the Android platform, one limitation of the study was that all participants were required to use an Android smartphone. A total of 15 families were initially recruited to participate in the study,

[1]The development screening test used focused on children of this particular age [43]

We excluded 2 families from our analysis because they did not use BebeCODE even from the beginning of the study. We tried to contact them to obtain an understanding of their lack of usage but we could not reach them. Another family did not even start due to personal circumstances. Thus, 12 families with children between the ages of 10 and 50 months participated in the study, and their information is summarized in Table 2. Each participant was compensated approximately $10 per week for participating in the study. Institutional Review Board (IRB) approval was obtained prior to the study from the university.

Table 2. Participant families and demographics (The ages in months of the children were based on their ages at the start of the experiment. Daycare is described in terms of hours per day.)

Family ID	Occupation (Father)	Age (Father)	Occupation (Mother)	Age (Mother)	# of children	Daycare	Child age	Child gender
F1	Office worker	30s	Housewife	20s	1	N/A	10m	F
F2	Student	30s	Housewife	30s	1	N/A	16m	M
F3	Public official	30s	Housewife	30s	1	N/A	38m	M
F4	Public official	30s	Housewife	30s	1	N/A	13m	M
F5	Office worker	40s	Housewife	20s	1	N/A	20m	F
F6	Carpenter	30s	Housewife	30s	1	6.5	50m	F
F7	Public official	30s	Housewife	30s	2	7	32m	M
F8	Public official	30s	Public official	30s	1	2	23m	F
F9	Student	30s	Student	30s	1	5	14m	F
F10	Office worker	30s	Student	30s	2	7	35m	M
F11	Public official	40s	Public official	30s	3	8	30m	F
F12	Office worker	30s	Office worker	30s	2	9	39m	F

Study Procedure

We uploaded BebeCODE on Google Play Store as a beta test version and sent the download URL to participants with a unique ID to connect to the server for loading their child's

information, which had been already registered in the system based on the pre-survey. Before starting the experiment, we published a manual explaining how to use BebeCODE and provided a short orientation over a phone to introduce the goal of the experiment and gave short instructions. Program usage patterns were recorded on the server for analysis such as the number of log-ins, the time of questions answered.

We conducted a 4-week BebeCODE deployment study with 12 sets of parents with children between 10 and 50 months of age. Because child development questions differ for children in different age groups, each set of parents answered the questions appropriate for their child's age group. However, during 4-week study, some participants (F4, F8, F9) answered that their child could do all the activities in the questions early in the study and that there were no questions about which there was disagreement. Since our primary goal was to observe the collaboration and consensus building between caregivers, we decided to give them an additional questionnaire after consulting with an expert (e.g., presenting questions for 36 to 41 months to the parents of a 34-month old child). After the 4-week period, we administered a web-based post-study survey included questions about the experience of using each feature such as collaboration of parents, notifications, and disagreement resolution, in addition to receiving feedback on how to improve the design of BebeCODE.

EXPERIMENTAL RESULTS

This section presents findings from the BebeCODE deployment study on how the participants used the system with their spouses, resolved disagreement, and tracked their children's development. All quotes in this section are translated from Korean.

Overall Usage Pattern

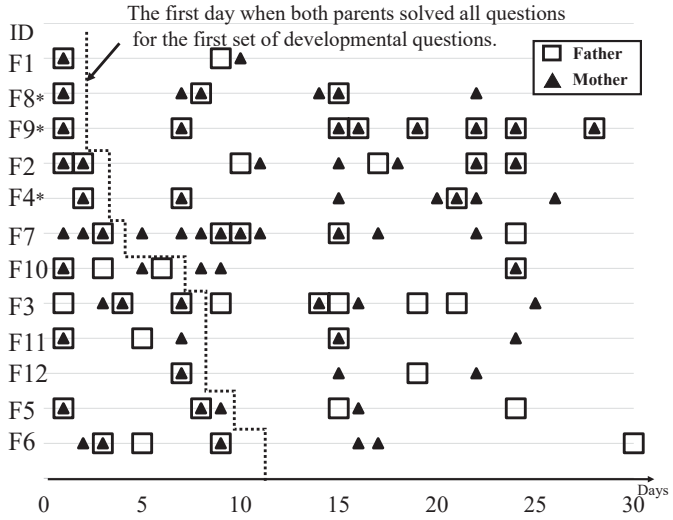

Figure 4. Overall usage pattern during 4-week deployment study. Each mark shows a day when a parent answered more than one question. (* notes participants who received the next set of developmental questions early.)

We examined the usage pattern of BebeCODE shown in Figure 4 via log files from the server to determine how often the parents accessed BebeCODE and answered the questions. The *X*-axis represents time or the duration of our study. Each mark indicates that more than one question had been answered. The dotted line shows the first day when both parents had answered all questions for the first set of developmental questions. All participant answered every question within 8 days (*M=4.5, SD=3.08*), and there were approximately 9.75 questions for which there was disagreement (*SD=7.3*) out of 45 per family. By the last day, 9 of 12 sets of parents resolved all of their disagreement to reach a consensus, but the rest of the parents had 1.3 unresolved disagreements on average (*SD=0.81*) because the parents could not test the questions that required specific items: "Can the child ride a bicycle with a secondary wheel?" or "Can the child stack blocks like a pyramid?" [43]. After resolving disagreements for all questions, they continuously accessed and fixed the answers. According to the post-survey, the parents reported that they fixed the answers when they resolved disagreement or when they received timeout notifications requiring them to re-check the questions regarding activities that their child could not do.

Essential Participation of Both Parents

Both parents had to participate in the assessment by responding to developmental questions. The participants could not see their spouses' responses until after they had answered. Most participants expressed that this design helped them not to be swayed by their spouse's response and to think about their child in depth for a while.

"Since the mother spends a lot of time with my child, if I saw her response, I might be easily swayed by her response." (*F2-F*[2])

"When I did not know the other person's opinion, I tried to think carefully about the development of my child." (F11-M)

Also, the final results could not be checked if there were any questions that had not been completed or about which there was disagreement. By inducing parents to reach a consensus on the disagreements, they had to communicate to answer the questions. We found positive effects of both parents' participation, which will be summarized below.

Objective assessment

The results of parent-report assessment depend on the respondent's accuracy, but a single parent cannot observe all of a child's activity. Because both parents have to participate in the assessment, they share their observations and experiences of their children, which results in more objective and accurate results.

"Even though I usually spend most of the time with my child, there were some things that I could not observe but the child's father could." (F2-M)

"I was not able to see if my child could fasten a button, but when I heard the story of my wife, I was able to better observe what my child was doing and be more objective." (F9-F)

Increased interest in children

BebeCODE encourages both parents to participate in development assessment, and it uses a timeout algorithm to notify

[2]We refer to each participant using following notation: [Family ID]-[Father or Mother]

parents to reconsider questions about activities which they answered that their child could not do. The participants were able to find out new things about their child that they were not aware of through answering the developmental questions with BebeCODE and address developmental questions with different answers. They reported that they started to observe their children more carefully than before to catch various actions and assess their development correctly.

"I did not consider my child's development carefully, but I started to observe my child more closely through the development questions" (F9-M)

"The father seemed to be more interested in watching the baby's behavior than before." (F4-M)

The increased interest in the development of their child led to parents being emotionally moved when they were able to confirm the growth and development of their child.

"I thought that the mother was exaggerating the child's activity, but a few days later, I saw the actual activity of the child. I was almost crying because I realized how fast my child was growing." (F4-F)

Improving communication between parents

The parents had to communicate to reach a consensus on questions that they disagreed on. They had to discuss their children's development and consider together how to help their children improve areas where improvement was needed. This improved the communication of parents to consider their children together.

"Our child's development is fast in most areas except the gross motor skills, so we discussed how we can help develop the child's gross motor skills. We agreed to reschedule our weekend so we can take walks more frequently." (F5-M)

"We had to share each other's thoughts, experiences, and evidence to answer the questions that we disagreed on." (F2-F)

Experience in Resolving Disagreements

Since there can be different opinions on developmental questions between parents, BebeCODE provides three features to support the resolution of disagreements through chatting, image/video sharing, and asking a third person. When the differences in opinion were resolved, BebeCODE asked the participants, through a pop-up message on the screen; *"which feature was helpful to resolve your disagreements."* The answers (N=115) selected were chatting/conversation (76%), image/video sharing (8%), asking a third person (10%), etc. (6%) (e.g., they had their children do or perform the specific activity to re-assess.). Interestingly, the results showed that simple chatting/conversation is helpful to reach a consensus in most cases. In the following sections, we explore in detail how the participants used each feature.

Chat

Chatting or instant messaging through the mobile system is the simplest way to communicate with others who are not nearby. Half of the participants used the chatting feature to resolve the questions that they disagreed on, but the others preferred

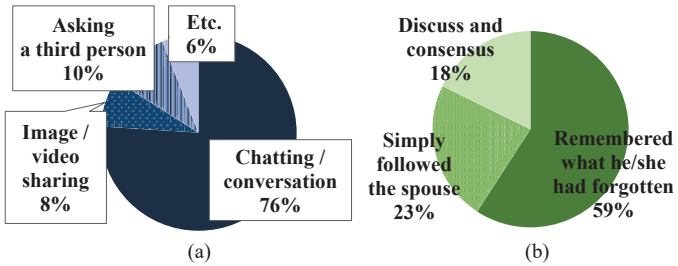

Figure 5. (a) Methods used to resolve the disagreements and (b) reasons why during the deployment study parents changed their opinion through chat or conversation

face-to-face conversation (i.e., they rarely used chatting) at home.

"When I disagreed, I could start a chat session and address the disagreement. If I send an instant message, my spouse checks it right away. It was easy to use and helped narrow difference of opinion." (F10-F)

"Face-to-face talking/discussing was better since I was able to communicate faster and received greater depth of feedback." (F8-M)

We asked each parent what made him/her change his/her opinion through chat/conversation since there was no physical evidence that their child could achieve the development question. Fifty-nine percent of the respondents reported that they remembered what they had forgotten through chatting/conversation with their spouses, and 23% of the answers simply followed the opinions of their spouses. The others responded that they reached a consensus on the questions for which there was not an immediately clear answer through chatting/conversation and discussion.

Picture/Video sharing

Through simple chatting/conversation, most of the disagreements were resolved. However, images and/or videos that showed children's specific activities further helped the parents since they provided strong evidence of the children's behavior. In addition, for parents who tended to simply agree with their spouses, the images and videos were helpful to solidify their decision.

"I was able to better start remembering some aspects of my child's behavior through past photos and videos taken by my spouse. One such example was seeing my child jump from the stair with both feet." (F9-M)

"There is no way to have better confidence that my recollection is accurate than to share pictures and videos." (F5-F)

Although the uploaded images were mostly used by parents to share with their spouses, as a side-effect, BebeCODE ended up providing a development-progress photo album that matched developmental questions with images. We asked the parents what were the advantages of uploading pictures and videos to BebeCODE. More than half the parents (64%) responded that they uploaded them to share with their spouses to reach a consensus, but some parents (32%) responded that it was simply to record their children's growth and create an "album."

They reported that a developmental album that shows children's progress has a different meaning compared to a normal picture album.

"There would be significant sentimental value to my child if we can record the development progress that he achieves." (F7-M)

Asking a third person

A third person who spends time with children can also answer some developmental questions [20]. Among the 12 families who participated in the study, 7 families had children that attended daycare center. Three of these families asked a third person, such as a daycare teacher, an uncle, or a friend to answer the questions. The main solution for one family was asking a daycare teacher. The mother reported that she asked a teacher questions when bringing her child to the daycare center and taking her child home.

"When I brought my child to the daycare center, I asked the teacher about what I was wondering (...) She was especially helpful because she had the most objective opinion among others." (F5-M)

Notifications for Continuous Tracking and Collaboration

BebeCODE sends two type of push notifications for parents. The first type of notification is to let parents know the responses, which is checked their child can't do, were disappeared since 7-days has passed. This type of notification was helpful to periodically track their children's development.

"When the message came, I wondered if my child could now do that particular activity [something he was not able to do earlier]. As the child succeeded, I was able to update my response to the development questions" (F10-M)

During the 4-week period, one parent was able to identify how the notification messages helped to identify growth in one particular area of their child.

"My child had difficulty playing shape games at first, but during the last week of the study, he was able to identify shapes easily. My husband and I were delighted!" (F4-M)

The other notification is sent when meaningful progress has been made by a parent's spouse, such as all questions in a section have been answered, a chat message or an image/video have been uploaded, or disagreement questions have been answered. These notifications were helpful to know how one's spouse is progressing.

"I was relieved to be able to confirm my husband's progress and was able to ask him to answer the question." (F5-M)

Other Findings and Suggestions for Improvement

We believe that the deployment study showed that BebeCODE can successfully encourage collaboration between parents to assess their children's development. All participants were able to continuously track their children's development through timeout notifications and resolve most of their disagreement to reach a consensus. During the survey conducted after the deployment study, 22 out of the 24 participants reported that they would be willing to continue to use BebeCODE, while the remaining 2 participants were also willing to continue using it if some aspects of the application interface could be improved. In addition, several comments on improving BebeCODE were provided that could be considered to help enhance BebeCODE.

Most parents of children who were showing signs of developmental delay were curious to know how they could address the developmental delay that their child might be facing. The initial BebeCODE system focused on encouraging collaboration between parents to their assess children's development, but it remains to be seen how the system can be extended to help provide treatment and/or intervention.

"I was wondering how to address the delayed physical development section. How do I support him? What is the solution?" (F7-M)

The developmental questions change every few months (e.g., 36-41 months, 42-47 months, 48-53 months ...). Even if a 48-month-old child is determined to be late in development, he/she has about 5 months to grow more until 53 months. In that case, the child is not actually considered to be experiencing delayed development, so the result message should refer to trying later.

"It's a developmental process for 21-23 months. I don't want to be frustrated by seeing the results in advance" (F8-M)

The developmental questions automatically change to next level as a child gets older. There were some children whose development was fast, such that the parents did not have questions that needed to be tracked because their children already could achieve all of developmental questions. In our study, we manually changed the developmental questions to the next stage when a child had an almost perfect score, and we enabled the parents to continuously track their child's development.

"My child's development was fast, so I had to wait until the questions changed to the next stage." (F2-F)

BebeCODE was improved so that the system would automatically change questions to the next stage based on a child's development.

DISCUSSION

Consensus Building as Primary Measure

In our evaluation, we chose not to directly compare parents' consensus against expert opinion. Parental development screening tests are intended for long-term, in-home observations, while experts mostly make observations during short sessions in a clinical setting. Because of such differences, simply comparing the two could be misleading since the two sources can provide complementary perspectives. The Denver Developmental Screening Test [13] includes parents' report of their child's behavior along with professional direct assessment. A recent study indicated that parent's opinions and expert assessment are both needed to obtain more reliable results [28]. Since both parents respond and their discussions and supporting materials (e.g., photos) are captured by our system, we see potential for experts to use this data in their assessments as well. In addition, we found positive evidence that shows better results of collaborative tracking based on the participant's reports.

Better recollection through communication: There were many instances in which parents resolved their disagreements with objective proof by sharing images/videos or by following a parent who made a direct observation or by remembering what they had forgotten through communication. Based on these observations, we expect more reliable answers to developmental questions.

Collect other's opinion: The parents discussed the developmental questions that they could not answer with certainty. Some of them asked a daycare teacher, a cousin, or a friend about their children's development. In such cases, the accuracy of others' opinions might be better or worse. However, at least, this could help parents reduce uncertainty in their assessments and make a more informed decisions.

Timeout Algorithm for Periodic Re-evaluation

Since children grow at a fast rate, frequent tracking of their development status is important. We applied a 7-day timeout algorithm, that *invalidated* questions after 7 days when the parents responded that the child *"can't do"* since the child's development could have changed. For these questions, BebeCODE sends push messages to ask parents to whether answer them to see if the child's status has changed. This algorithm can support not only response validity but also frequent tracking. We believe that the deployment study showed the timeout algorithm worked well based on positive feedback from participants. However, the timeout parameter should change depending on age as the development of children varies [20]. During our 4-week experiment, we set up the timeout interval to be 7 days for all children, and on average, 9.3 (SD=8.01) timeout notifications were sent to each participant. Weekly notification might be annoying to some parents, as one participant said: *"There were many cases in which there was not much change in the child's development compared to the previous week. For my child, one week was too short for the timeout algorithm (F6-father)."*

Diverse Family Considerations

Our current system design/evaluation did not address diverse family structures as we limited the scope of our scope study to families with a mother and a father sharing the same household. We believe that BebeCODE can be extended to support diverse families, including single parents, more than two caregivers, etc. One extension would be to allow customization of caregiver settings, in terms of the number of primary caregivers (e.g., 1 for single parents; more than 2 if other family members such as grandparents are involved) and their roles (e.g., primary group that would be required to answer the questions and reach consensus; optional group who can be contacted for opinions but input or consensus is not required).

Extension to Other Assessments

In this paper, we applied our system design for the assessment of child development. However, the design can be extended to other observation-based assessments that have the limitations of depending on respondent's observation or memories. There are many kinds of assessment based on parent reports in a clinical context, such as Ages and Stages Questionnaires [5] and Child Development Inventories [20] for screening children development or Brief Infant-Toddler Social and Emotional Assessment [17] and Behavioral Assessment of Baby's Emotional and Social Style [12] for a social-emotional screening test. Thus, we believe that our design will make a valuable contribution that can be extended to other types of assessment to encourage collaboration between parents and caregivers.

Limitations

Our 4-week deployment study with 12 families was likely too short to confirm whether participants could track children's long-term development. Fortunately, there were three children who were approximately 12 months old, which is an age at which a child's development changes relatively quickly [43, 20]. From our post-interview, one parent was able to track their child's growth through BebeCODE, rather than just checking the current child's status, so we believe that our system can help parents track children's development if it is used for a longer period of time. Even though BebeCODE supports collaboration between parents and other caregivers, there were only 7 participants who attended a daycare center, and only one family frequently asked the daycare teacher about uncertain aspects of the child's development progress, which shows that parents can get useful information from other caregivers. It remains to be seen how BebeCODE can be extended to encourage participation of other caregivers.

SUMMARY

In this paper, we presented the design and implementation of BebeCODE, a collaborative mobile system to assess child development. BebeCODE encourages parents to answer all developmental questions independently and resolve disagreements to reach a consensus via chatting, image/video sharing, and asking a third person. We conducted a 4-week deployment study and found that BebeCODE successfully helped parents participate in their children's developmental assessment and resolve their disagreement to reach a consensus.

ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers and KIXLAB members for their valuable comments and feedback. This research was also supported in part by National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (MSIP) (2015M3C4A7065647) and (2017R1A2B4011457) and in part by an Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (No.2017-0-00537, Development of Autonomous IoT Collaboration Framework for Space Intelligence).

REFERENCES

1. Craig A. Albers and Adam J. Grieve. 2007. Test Review: Bayley, N. (2006). Bayley Scales of Infant and Toddler Development Third Edition. San Antonio, TX: Harcourt Assessment. *Journal of Psychoeducational Assessment* 25, 2 (2007), 180–190.
2. William Barnett. 2011. Effectiveness of Early Educational Intervention. *Science (New York, N.Y.)* 333 (08 2011), 975–8.
3. L. Boccanfuso, E. Barney, C. Foster, Y. A. Ahn, K. Chawarska, B. Scassellati, and F. Shic. 2016. Emotional

robot to examine different play patterns and affective responses of children with and without ASD. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 19–26.

4. Yeargin-Allsopp M. Boyle C.A., Decoufle P. 1994. Prevalence and health impact of developmental disabilities in US children. *Pediatrics,* 93, 3 (Mar 1994), 399–403.
5. Diane Bricker and Jane Squires. 1999. *Ages and Stage Questionnaires A Parent-Completed, Child-Monitoring System*. Paul H. Brookes, Maryland.
6. Glendon Casto and Margo A Mastropieri. 1986. The efficacy of early intervention programs: A meta-analysis. *Exceptional children* 52, 5 (1986), 417–424.
7. F. Cecchi, S. M. Serio, M. Del Maestro, C. Laschi, and P. Dario. 2010. Design and development of sensorized toys for monitoring infants' grasping actions. In *2010 3rd IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics*. 247–252.
8. Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A. Kientz, and Sean A. Munson. 2016. Boundary Negotiating Artifacts in Personal Informatics: Patient-Provider Collaboration with Patient-Generated Data. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 770–786.
9. Daniel T. Williams Clarice J. Kestenbaum. 1988. *Handbook of Clinical Assessment of Children and Adolescents*. New York University Press.
10. Stephan Collishaw, Robert Goodman, Tamsin Ford, Sophia Rabe-Hesketh, and Andrew Pickles. 2009. How far are associations between child, family and community factors and child psychopathology informant-specific and informant-general? *Journal of Child Psychology and Psychiatry* 50, 5 (2009), 571–580.
11. Munson J. Smith M. Winter J. Greenson J. Dawson G., Rogers S. 2010. Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics* 125 (2010), 17–23.
12. Karen Moran Finello and Marie Kanne Poulsen. 1996. The behavioral assessment of baby's emotional and social style (babes): A new screening tool for clinical use. *Infant Behavior and Development* 19, Supplement 1 (1996), 453.
13. William K. Frankenburg and Josiah B. Dodds. 1967. The Denver Developmental Screening Test. *The Journal of Pediatrics* 71, 2 (1967), 181 – 191.
14. Stevenson C.S. Lewit E.M. Behrman R.E. Gomby D.S., Larner M.B. 1995. Long-term outcomes of early childhood programs: analysis and recommendations. *Future Child* 5, 3 (1995), 6–24.
15. Elisa Hamer, Tjitske Hielkema, Arend Bos, Tineke Dirks, Siebrigje J. Hooijsma, Heleen Reinders-Messelink, Rivka F. Toonen, and Mijna Hadders-Algra. 2017. Effect of early intervention on functional outcome at school age: Follow-up and process evaluation of a randomised controlled trial in infants at risk. *Early Human Development* 106-107 (03 2017), 67–74.
16. Matthew K. Hong, Lauren Wilcox, Daniel Machado, Thomas A. Olson, and Stephen F. Simoneaux. 2016. Care Partnerships: Toward Technology to Support Teens' Participation in Their Health Care. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5337–5349.
17. Gabriela M. Hungerford, Dainelys Garcia, and Daniel M. Bagner. 2015. Psychometric Evaluation of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) in a Predominately Hispanic, Low-Income Sample. *Journal of Psychopathology and Behavioral Assessment* 37, 3 (01 Sep 2015), 493–503.
18. J. McV. Hunt and John Paraskevopoulos. 1980. Children's Psychological Development as a Function of the Inaccuracy of Their Mothers' Knowledge of Their Abilities. *The Journal of Genetic Psychology* 136, 2 (1980), 285–298.
19. Inseok Hwang, Chungkuk Yoo, Chanyou Hwang, Dongsun Yim, Youngki Lee, Chulhong Min, John Kim, and Junehwa Song. 2014. TalkBetter: Family-driven Mobile Intervention Care for Children with Language Delay. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 1283–1296.
20. Harold Ireton. 1992. *Child Development Inventory Manual*. Behavior Science Systems, Minnesota.
21. Aesha John, Am Halliburton, and Jeremy Humphrey. 2013. Child mother and child father play interaction patterns with preschoolers. *Early Child Development and Care* 183, 3-4 (2013), 483–497.
22. Bumsoo Kang, Chulhong Min, Wonjung Kim, Inseok Hwang, Chunjong Park, Seungchul Lee, Sung-Ju Lee, and Junehwa Song. 2017. Zaturi: We Put Together the 25th Hour for You. Create a Book for Your Baby. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1850–1863.
23. Kidsnote. 2017. `https://play.google.com/store/apps/details?id=com.vaultmicro.kidsnote&hl=ko`. (2017).
24. Julie A. Kientz, Rosa I. Arriaga, Marshini Chetty, Gillian R. Hayes, Jahmeilah Richardson, Shwetak N. Patel, and Gregory D. Abowd. 2007. Grow and know: Understanding record-keeping needs for tracking the development of young children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1351–1360.

25. Julie A. Kientz, Rosa I. Arriaga, and Gregory D. Abowd. 2009. Baby Steps: Evaluation of a System to Support Record-keeping for Parents of Young Children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1713–1722.

26. P.W. Mattessich, B.R. Monsey, Wilder Research Center, and Amherst H. Wilder Foundation. 1992. *Collaboration–what makes it work: a review of research literature on factors influencing successful collaboration*. Amherst H. Wilder Foundation. https://books.google.co.jp/books?id=_RZHAAAAMAAJ

27. Andrew D. Miller, Sonali R. Mishra, Logan Kendall, Shefali Haldar, Ari H. Pollack, and Wanda Pratt. 2016. Partners in Care: Design Considerations for Caregivers and Patients During a Hospital Stay. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 756–769.

28. Lauren E. Miller, Kayla A. Perkins, Yael G. Dai, and Deborah A. Fein. 2017. Comparison of parent report and direct assessment of child skills in toddlers. *Research in Autism Spectrum Disorders* 41-42, Supplement C (2017), 57 – 65.

29. Francisco Nunes and Geraldine Fitzpatrick. 2015. Self-Care Technologies and Collaboration. *International Journal of Human–Computer Interaction* 31, 12 (2015), 869–881.

30. OECD. 2015. How's Life? (2015).

31. Laura R. Pina, Sang-Wha Sien, Teresa Ward, Jason C. Yip, Sean A. Munson, James Fogarty, and Julie A. Kientz. 2017. From Personal Informatics to Family Informatics: Understanding Family Practices Around Health Monitoring. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 2300–2315.

32. Ari H. Pollack, Uba Backonja, Andrew D. Miller, Sonali R. Mishra, Maher Khelifi, Logan Kendall, and Wanda Pratt. 2016. Closing the Gap: Supporting Patients' Transition to Self-Management After Hospitalization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5324–5336.

33. Katherine Van Dusen Randazzo, John Landsverk, and William Ganger. 2003. Three Informants' Reports of Child Behavior: Parents, Teachers, and Foster Parents. *Journal of the American Academy of Child Adolescent Psychiatry* 42, 11 (2003), 1343 – 1350.

34. Rosa Rosnati, Rosario Montirosso, and Daniela Barni. 2008. Behavioral and Emotional Problems Among Italian International Adoptees and Non-Adopted Children: Father's and Mother's Reports. *Journal of the Division of Family Psychology of the American Psychological Association* 22 (09 2008), 541–9.

35. Von Suchodoletz W. Sachse S. 2008. Early identification of language delay by direct language assessment or parent report. *Journal of Developmental and behavioral pediatrics* 29, 1 (2008), 34–41.

36. M. M. Serrano, Y. P. Chen, A. Howard, and P. A. Vela. 2016. Lower limb pose estimation for monitoring the kicking patterns of infants. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2157–2160.

37. Seokwoo Song, Juho Kim, and John Kim. 2017. PlayBetter: A Phone-based Baby Play Support System for Childcare Bystander Parents. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2100–2106.

38. Seokwoo Song, Seungho Kim, John Kim, Wonjeong Park, and Dongsun Yim. 2016. TalkLIME: Mobile System Intervention to Improve Parent-child Interaction for Children with Language Delay. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 304–315.

39. Hyewon Suh, John R. Porter, Alexis Hiniker, and Julie A. Kientz. 2014. @BabySteps: Design and Evaluation of a System for Using Twitter for Tracking Children's Developmental Milestones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2279–2288.

40. Hyewon Suh, John R. Porter, Robert Racadio, Yi-Chen Sung, and Julie Kientz. 2017. Baby Steps Text: Feasibility Study of an SMS–Based Tool for Tracking Children's Developmental Progress. *AMIA Annual Symposium Proceedings* 2016 (02 2017), 1997–2006.

41. Tammy Toscos, Kay Connelly, and Yvonne Rogers. 2012. Best Intentions: Health Monitoring Technology and Children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1431–1440.

42. Tracy L. Westeyn, Gregory D. Abowd, Thad E. Starner, Jeremy M. Johnson, Peter W. Presti, and Kimberly A. Weaver. 2012. Monitoring Children's Developmental Progress Using Augmented Toys and Activity Recognition. *Personal Ubiquitous Comput.* 16, 2 (Feb. 2012), 169–191.

43. Chung-Hyuk Yim, Gun-Ha Kim, and Baik-Lin Eun. 2017. Usefulness of the Korean Developmental Screening Test for infants and children for the evaluation of developmental delay in Korean infants and children: a single-center study. *Korean journal of pediatrics* 60, 10 (October 2017), 312–319.