Training, Education, and Learning Sciences, Vol. 193, 2025, 75–83
https://doi.org/10.54941/ahfe1006648

Towards Co-Design Workshops Based on Data-Driven Context Detection: A Pilot Reflection Workshop in Childcare

Yuki Taoka[1], Sawako Fujita[1], Shigeru Owada[2], Shiori Fujimaki[1], Kaho Kagohashi[1], Momoko Nakatani[1], and Shigeki Saito[1]

[1]Institute of Science Tokyo, Tokyo, Japan
[2]Sony Computer Science Laboratory, Tokyo, Japan

ABSTRACT

To improve childcare quality, co-design involving both designers and practitioners is essential. This study explores the impact of data-driven scene extraction on reflection in co-design workshops. Video recordings were collected in a Japanese nursery through self-logging and automatic sound-based selection. In a workshop with eight childcare workers, participants reflected on both types of videos, analyzing behaviors and insights. The results show that behavioral data and automated video excerpts support reflection effectively. Future work will further examine the role of behavioral data in co-design, involving both daycare staff and designers to enhance childcare practices.

Keywords: Data-driven reflection, Behavioral data, Co-design

INTRODUCTION

Improving the quality of childcare is a crucial issue, as it plays a significant role in children's identity formation. The OECD defines the service quality of Early Childhood Education and Care (ECEC) as "the features of children's environments and experiences provided by an ECEC setting that contribute to children's well-being" (OECD, 2015). To enhance the quality of childcare, it is essential to identify current challenges—namely, to reflect deeply on and understand childcare practices and then develop solutions based on these insights. These solutions should not only focus on improving childcare providers' behaviors but also on enhancing the environment through the development of new products and services for childcare facilities.

To address these challenges, co-creation with diverse stakeholders, including childcare professionals, experts in childcare environments, and service designers, can be an effective approach. This process, known as co-design, involves collaboration between designers and non-designers who have not received formal design education (Sanders & Stappers, 2008). For co-design to be effective, both designers and non-designers must develop a deep understanding of what is happening in the field (Pirinen & Others, 2016).

In the childcare field, much of the practical knowledge is context-dependent and can only be fully understood by those working on-site. Childcare providers, as practitioners striving daily to improve the quality of childcare, can offer detailed insights into specific situations in the field. Experts, on the other hand, can leverage their specialized knowledge to identify new issues that may not be apparent to practitioners. In other words, in co-design, the starting point is to gain the deepest and most concrete possible understanding of the practical realities of childcare settings.

One widely used research method for gaining a deeper understanding of specific situations is the direct observation of childcare providers in their actual work environment. This approach is effective as it allows designers to capture real-time details of childcare situations and challenges. However, direct observation in childcare facilities may disrupt daily operations, and stakeholders often face restricted access to the field. Therefore, when direct observation is not feasible, alternative methods that utilize existing materials from the facility become necessary.

In such cases, the design research community has proposed various tools and methods to facilitate an understanding of the field (Sanders et al., 2010). Examples include written diaries (Steen et al., 2011), video observations (Lee et al., 2018), and video diaries (Rose & Cardinal, 2018). Among these methods, video-based approaches have proven to be highly effective. For example, in the video diary method, co-designers record videos in response to specific questions, allowing for a structured yet flexible way of capturing their experiences. Additionally, various co-design methods, such as cultural probes and generative toolkits, are used to explore participants' values (Sanders & Stappers, 2012). Cultural probes involve taking photographs of specific objects, while generative toolkits enable participants to express latent values through hands-on activities.

With recent advancements in information and communication technologies (ICT), it has become possible to collect in-situ behavioral data. While such data has not yet been widely adopted in design research, its potential to enhance co-design by providing objective and context-rich insights is significant. Using data-driven tools for reflection enables stakeholders to leverage field-generated data. These materials provide a shared understanding of specific childcare scenes, allowing co-design participants to reflect on and analyze practices more effectively. In particular, video data serves as an effective tool for objective reflection on childcare practices and fostering a shared understanding among co-design participants.

This allows for continuous data collection without disrupting daily practices. However, since childcare activities take place from morning to night every day, manually reviewing vast amounts of recorded video is impractical. Therefore, a mechanism for efficiently extracting meaningful scenes is needed.

One approach is **self-logging**, where childcare providers actively annotate key moments during their practice. This method ensures that the recorded data reflects practitioners' perspectives but requires their direct input, which may be burdensome. An alternative approach is **automatic video extraction**, which records all activities and later identifies meaningful clips without

requiring manual annotation. While this method reduces the workload on practitioners, its effectiveness in facilitating co-design remains uncertain.

This study compares self-logging and automatic video extraction to evaluate their effectiveness in supporting co-design in childcare settings. By examining these approaches, we aim to clarify their respective strengths and implications for collaborative design.

SELF-LOGGING AND AUTOMATICALLY EXTRACTED VIDEO

In the co-design process aimed at improving the quality of childcare, we adopt video recordings as a tool for understanding childcare settings. As mentioned in the previous section, we employ two types of video recordings: self-logging and automatic video extraction. The following section provides a detailed explanation of each method.

Self-Logging

For self-logging, this study used VisRef (Owada et al., 2023), a smartwatch application that works in coordination with a smartphone to allow participants to record childcare workers' insights during practice. With VisRef, users can start recording with a simple touch on the smartwatch, making annotation easy and accessible. This method is expected to incorporate practitioners' perspectives into co-design while minimizing their workload.

Childcare practitioners do not act solely based on experience; rather, they engage in continuous reflection while immersed in their practice, as Schön (1983) describes in his concept of reflection-in-action. These reflections contain a wealth of practical knowledge that is constantly shaped and refined through experience. Encouraging childcare workers to engage in self-logging during their activities is expected to help externalize this implicit knowledge and further enhance their reflective processes.

Annotations can be made using the smartwatch's voice recording function or by selecting predefined sentences on the smartphone screen. Upon a self-logging request, the system records approximately two minutes before and 30 seconds after the request, resulting in a three-minute video segment. This setup allows for real-time annotation with minimal disruption to daily activities.

Automatic Video Extraction

For automated extraction, we focus on audio volume, which can highlight noteworthy moments, such as when children cry or engage in highly energetic interactions. Audio-based extraction has been applied in various fields, including media analysis, where it helps identify key scenes in news or movies (Sidiropoulos et al., 2011), and sports broadcasting, where commentators' voice intensity is used to detect peak moments (Otsuka et al., 2005). In pediatric medicine, children's crying sounds have also been automatically extracted for medical analysis (Cabon et al., 2018).

This method may also be useful for extracting meaningful data from childcare videos to enhance field understanding. A rise in audio volume in

a childcare setting often indicates significant events that prompt children or childcare providers to raise their voices. Therefore, simple extraction based on sound levels could serve as a valuable cue for scene selection.

To verify the concept of automatic detection based on sound levels in an ECEC setting, we employed the Wizard of Oz method. One of the authors manually used video editing software to extract three-minute video segments—1.5 minutes before and after the moment when the volume exceeded a predefined threshold. Next, scenes in which no children were present and where there was nothing relevant to discuss were removed, such as footage of a childcare worker vacuuming in an empty room. Furthermore, to eliminate false positives—such as temporary volume spikes caused by objects falling or furniture being moved—we prioritized videos with the highest total volume levels within the selected three-minute window.

DATA COLLECTION AND ANALYSIS

We conducted a workshop with eight childcare workers who work at a childcare facility in a metropolitan area in Japan. The video was recorded on a single day in a nursery room used mainly for one-year-old children. The room is also used for mixed-age childcare outside of regular childcare hours. A chief staff member taking care of one-year-old children used the self-logging method. In the workshop, the participants watched two self-logged videos and two automatically detected videos to compare the differences due to the recording methods. Table 1 briefly explains the four videos. The participants were divided into groups of two or three. One of the participants in each group cared for children during the data collection so that the member could provide contextual information on the childcare of the day. The other members were distributed to create a diversity of experience and daily responsibilities. Each group was facilitated by a researcher. The researcher also supported the manipulation of the iPad to watch the recordings.

The workshop lasted 2 hours and was divided into two steps. Before starting the workshops, we emphasized that reflection in a workshop is not to conclude at a single correct answer but to share each person's thoughts and gain new insights. After a quick icebreak session, the participants watched the four selected videos together. The two self-logged videos were explained by the person who selected them. The explanation includes the intention behind the selection and the childcare context. The two automatically selected videos were shown to the participants without the explanations. The participants were encouraged to take note of what they noticed by watching the videos. After having watched the four videos, the participants shared and discussed their findings within each group. The participants could watch the videos with the iPad placed at each table. The order of watching videos was counterbalanced to remove the order effect. Participants first watched their self-logged video and then viewed an automatically selected video. After that, the participants watched another self-logged video, followed by another automatically selected video. This first step lasted about 45 minutes. In the second step, the participants shared their discussion across the group, and they also discussed how the participants would like to reflect on their practice in the future. The second step lasted 35 minutes. All process was video recorded.

Table 1: The four videos used for the workshops.

Method	Video Summary
Self-logged	Timestamp: 9:29, Tag: Play Scene: The recorder was concerned about the situation in the nursery room without her presence. Upon reviewing, she noticed a child placing his hands on a box, nearly causing a shelf to fall, which alerted her to a hazardous environment.
	Timestamp: 11:10, Tag: Play Scene: The video was selected because the positioning and roles of the assigned caregiver were clearly visible in the video.
Audio volume	Timestamp: 8:28: A scene from a joint session where children from the 1-year-old class and other classes gathered together.
Audio volume	Timestamp: 8:28: A scene from a joint session where children from the 1-year-old class and other classes gathered together.

The semi-structured individual online interviews were conducted to assess the participant's perceived difference on the different selection methods and explore additional information that could have improved the discussion. The questions include "Did you notice any differences depending on the video you watched?" and "What additional information would you have wanted?". The interview lasted 20 minutes to 1 hour, depending on the participants' availability. The interviews were recorded and transcribed. The transcripts were labeled and organized into categories based on the similarity of the labels.

RESULT AND DISCUSSION

The Influence of the Different Cutting Methods

Table 2 presents the main insights of the participants during discussions using videos extracted by each method. In both extraction methods, participants discussed the childcare room environment and children's behavior. However, the placement of childcare workers was discussed only in the videos extracted using the self-logging method. It should be noted that the two automatically extracted videos were recorded during mixed-age childcare, while the self-logging recordings were taken during regular childcare hours. The purpose of childcare varied depending on contextual factors, such as the number of staff and children in the room. This difference may have influenced whether or not the placement of childcare workers became a topic of discussion.

Table 3 presents the results of the post-interview analysis, providing insights into the participants' impressions of the significance of reflection on scenes selected based on high audio volume. The high volume within the childcare room implies that the child is restless, and it was possible to reflect on situations where there was a potential risk of injury or where the nursery staff was not able to keep up with the situation. On the other hand, the self-logging method was useful in inferring the value of the childcare workers who selected the videos.

Table 2: Insights gained from videos with different editing methods.

Method	Summary of Insights Gained	Example Statements
Audio Volume	Environment during joint activity time	During joint activity time, there were moments that were not developmentally appropriate or lacked calmness, so I thought we should review this aspect.
	Children's behavior during joint activity time	It was a bit noisy during joint activity time. Since children of different ages were present, I noticed potential dangers when children from different age groups entered the one-year-old classroom.
Self-logged	Placement of caregivers	I realized that caregivers were not always positioned where they needed to be. Although I was aware of it, there were situations where we couldn't provide adequate care due to the number of children. I wanted to interact more attentively with each child, but it was realistically difficult due to the staff-to-child ratio.
	Children's behavior	The one-year-old children, being accustomed to daily care, seemed relatively calm.
	Environmental setup	The environment was being used effectively, but at the same time, I wondered if the materials placed were developmentally appropriate. There were also challenges related to how the corners were sectioned off.

Table 3: Significance of reviewing scenes extracted based on audio peak.

Method	Summary	Example Statements
Audio Volume	Reviewing the current situation (1 person)	Even when told that the scene had a loud noise, I just accepted it as usual. That child tends to jump around. However, when I thought about the child's needs, I realized that I hadn't been fulfilling them properly. I felt that I was overlooking the current situation.
	Observing moments of restlessness (2 people)	In daily childcare, there are times when children are restless. I felt that using audio volume to extract such moments could provide useful insights.
	Reducing the risk of injury (1 person)	I also believed that when the noise level is high in childcare settings, the risk of injuries increases. Reviewing moments with high noise levels could help reduce injury risks.

Continued

Table 3: Continued

Method	Summary	Example Statements
	Reflecting on joint activity time (4 people)	The way joint activities are conducted requires careful attention. How can we create an environment where everyone feels comfortable? That is a constant challenge. Children's interests change over time, and if only predetermined activities are provided, they become bored. I think this is connected to noise levels as well.
Self-logging	Observing caregiver growth (1 person)	I realized the significance of staff members being able to choose a specific moment from hours of childcare footage. Watching the video, I felt that my perspective had shifted significantly.

The results suggest no significant difference in the perspectives of the participants' perceptions in either of the two methods of extraction, showing that the video extracted at the volume peak could also be used for the reflection. On the other hand, one of the videos selected by the caregivers was a scene that was not loud but posed a risk of injury. Such videos cannot be cut out by the volume threshold. Therefore, it was shown that there is a possible agenda by extracting the video from the childcare worker's observations. The results imply that the two video cutout methods could work as complementary to each other.

Additional Information Required When Viewing Videos

When viewing the videos, there were four main types of information that the child caregivers wanted to know outside of the videos: The first was "the child's voice and the voice of the child caregiver. The second is the "background of the scene," which includes the child's personality and recent activities of the child. It was pointed out that this is useful for a deeper understanding of the situation of the children in the video. Regarding the "number of people" and "time of day," some respondents wanted to know the ratio of the number of caregivers to children and the time of day when the videos were taken in order to understand what the situation was like in the nursery.

The children's voices and the caregivers' voice calls cannot be heard simply by recording the entire nursery. However, as discussed in the PCIT and other publications, voice calls are considered important because they play an important role in the care of the children. Similarly, the child's personality is an important factor in determining the method of care and is considered necessary when watching videos of children who are not in contact during normal childcare hours.

In addition, the time of day also helps to understand the situation, as it is possible to infer the events taking place at different times of the day for childcare services. For example, since children's behavior differs depending on whether it is before or after lunch, it is considered effective to indicate

these time periods. The number of childcare workers could also be used as an index to assist in understanding how many childcare workers are in the room and how they are behaving, since there is a possibility that there are childcare workers in places not captured by the camera due to blind spots.

CONCLUSION

This study compared two video extraction methods by conducting a workshop with childcare workers. The results showed that both methods enabled discussion and allowed the participants to address new findings on the videos. The results suggest that the two methods can be used complementary to explore diversified topics. The necessary additional information may be potentially useful for the workshops, with video recordings of the practice in contexts other than childcare. Overall, this study provides preliminary findings on using videos for reflection.

ACKNOWLEDGMENT

This study was funded by JST-Mirai Program (grant number JPMJMI22H3).

REFERENCES

Cabon, S. et al., 2019. Video and audio processing in paediatrics: A review. *Physiol Meas*. 40(2), 02TR02.

Cherrington, S., Loveridge, J., 2014. Using video to promote early childhood teachers' thinking and reflection, *Teaching and Teacher Education*, 41, pp. 42–51.

Lee, J.-J. et al., 2018. Design choices framework for co-creation projects. *International Journal of Design*, 12(2).

OECD, 2015. *Starting Strong IV: Monitoring in Early Childhood Education and Care*.

Otsuka, I., et al. 2005. A highlight scene detection and video summarization system using audio feature for a personal video recorder. In IEEE Transactions on Consumer Electronics. 51(1), pp. 112–116.

Owada, S. et al., 2023. VisRef: A reflection support system using a fixed-point camera and a smartwatch for childcare fields. In *Lecture Notes in Computer Science*. Lecture notes in computer science. Cham: Springer Nature Switzerland, pp. 146–166.

Pirinen, A. & Others, 2016. The barriers and enablers of co-design for services. *International Journal of Design*, 10(3), pp. 27–42.

Rose, E. & Cardinal, A., 2018. Participatory video methods in UX: Sharing power with users to gain insights into everyday life. *Commun. Des. Q. Rev*, 6(2), pp. 9–20.

Sanders, E. B.-N., Brandt, E. & Binder, T., 2010. A framework for organizing the tools and techniques of participatory design. In *Proceedings of the 11th Biennial Participatory Design Conference*. 11th Biennial Participatory Design Conference. PDC '10: The 11th Biennial Participatory Design Conference. New York, NY, USA: ACM. Available at: https://dx.doi.org/10.1145/1900441.1900476.

Sanders, E. B.-N. & Stappers, P. J., 2008. Co-creation and the new landscapes of design. *CoDesign*, 4(1), pp. 5–18.

Sanders, E. B. N. & Stappers, P. J., 2012. *Convivial toolbox: Generative research for the front end of design*, Amsterdam, Netherlands: BIS Publishers B. V.

Schön, D., 1983. The Reflective Practitioner: How Professionals Think in Action. Basic Books, New York, USA.

Sidiropoulos, P., et al. 2011, Temporal Video Segmentation to Scenes Using High-Level Audiovisual Features, in *IEEE Transactions on Circuits and Systems for Video Technology*, 21(8), pp. 1163–1177.

Steen, M., Manschot, M. & Koning, N. D., 2011. Benefits of Co-design in Service Design Projects. *International Journal of Design*, 5(2), pp. 53–61.