Making a difference in the real world. User-centred impact evaluation of an eight-country, community-based early childhood programme

Lake, G, Urban, M, Giblin, F, French, G & Farrell, T

Abstract

This paper presents a unique approach to the Impact Evaluation of a project that focused on low-threshold intergenerational play-based interactions in order to support young children from marginalised communities in eight European countries. The approach builds upon the work of Fetterman's *Empowerment Evaluation* and Patton's *Utilization Focused Evaluation* and brings them together to form an adapted model of evaluation. We outline in this paper how these two well developed methods of evaluation have been applied to a real world context, that is, the impact evaluation of a complex international project. Our approach highlights the complexities of differing contexts and allows for surprising and unintended consequences to emerge. It results, through double loop learning, a type of feedback loop with the internal stakeholders and implementers that is useful to the project coordination team, with a view to further upscaling of the initiative. Recommendations for policy at local, national and European Union levels were provided to the project and potential external users. However, the predominant feedback was provided at two crucial points along the way; during a stakeholder mapping exercise and during the further development of monitoring data tools.

Introduction

This paper reports and discusses experiences with, and findings of, an impact evaluation of *TOY to Share, Play to Care*, which was a co-funded project by the European Union Erasmus+ Programme (Key Action 3- Social inclusion and common values) and Open Society Foundations within the intergenerational *Together Old and Young* (TOY) programme. *TOY to Share, Play to Care* was the

second phase of the *TOY for Inclusion* project (http://www.toyproject.net/project/toy-inclusion-2/) which provides inclusive early childhood education and care (ECEC) play hubs in marginalised communities in eight European countries (Croatia, Hungary, Italy, Latvia, The Netherlands, Slovakia, Slovenia and Turkey). The collaborative project which involves national partner organisations supporting sixteen community-based Local Action Teams, is managed by the international *TOY* programme coordinators, International Child Development Initiatives (https://icdi.nl/). A detailed impact evaluation report of the *TOY to Share, Play to Care* project has been produced (Authors, 2021) with summary evaluation reports in eight languages available (https://reyn.eu/impact-evaluation-and-policy-recommendations-executive-summary/). The focus of this paper is on the design, conduct and *lessons learned* of the impact evaluation of this project, and the implications of the application of user-centred evaluation in complex projects involving diverse policy and practice audiences.

Evaluation methodology frequently focuses on Theory of Change models (Ofek, 2017, Patton, 2015). These have been criticised for being too linear a process and dealing with outcomes that are concluded when the project is completed. These models provide little support to the project team as they implement it and the context within which it has been conducted (Cousins, Whitmore & Shula 2012). More recently, a more active approach to evaluation, which takes place during the project roll-out, has emerged. *Empowerment Evaluation* (EE)*, originally developed in the US by David Fetterman (Fetterman, Kaftarian, & Wandersman, 1996, 2015), facilitates evaluators from within the project to engage in high-level self-reflection in order to gain understanding of *what is going on* inside the programme from the users' perspectives. A team of external overseers facilitate learning while the project takes place. They provide opportunities for the implementation team to explore the values of the programme and their potential impact. The impact evaluator acts very much in a coaching role, as opposed to an external process evaluator in its truest sense. It has been used widely by agents including the *National Aeronautics and Space Agency* and the *US Department of Education*, and in diverse context including community health initiatives in South African townships, Native American peoples and schools in academic distress (Fetterman, 2001, 2013; Fetterman, Kaftarian, & Wandersman, 2015; Fetterman, Rodríguez-Campos, & Zukoski, 2018; Fetterman & Wandersman, 2005). These type of evaluations have predominantly been used for evaluating cross-cultural studies which are participatory in nature and are collaborative. They have been subject to much debate, as to the true levels of control the participants have in the evaluation, vis à vis the external team who oversee it. Fetterman (2013) argues that stakeholders are largely involved when using EE and that empowerment occurs at the individual level; in the shape of capacity building in decision-making and interpersonal skills and at the organisation level; where the responsibility for these decisions is shared (Miller, 2006). Furthermore, the three steps of EE, Fetterman contends, allow the evaluator and programme participants to be equals in:

1. establishing the mission statement;

2. taking stock of their circumstances, and;

3. agreeing goals for moving forward (Fetterman et al., 1996)

Criticisms levelled at the movement are centred around the use and misuse of terms associated with the practice. A large-scale review conducted by Miller in 2006 concluded that as a result of disagreement of users on the term 'empowerment', adherence to the true sense of Empowerment

Evaluation is problematic. As a result, there can be weaker outcomes in relation to true empowerment of the participants in these studies (Miller, 2006). Miller called for further studies to explore EE's use in wide-varying contexts. Fetterman and Wandersman (2007) responded to this article highlighting the actual omission of several current key international examples at that time. These included some key evaluation reports.

A later work by Scriven (2017) has highlighted a number of EE's still outstanding positive attributes. Namely, Scriven supports that the evaluation starts from within the project, i.e. the users of the programme. By doing this, it allows the implementers make changes to what they are implementing and is live in this process. Implementers can really get to know their programme which can, in turn, maximise their buy-in. Scriven does, however, question the validity of the model, as there is a proneness to bias (Scriven, 2017). He argues that this can be offset by the external team's involvement and their potential for objectivity.   It could be argued, however, that the application of all the principles of EE in their purest sense is difficult in a real world context. This is often due to evaluators being absent from the onset of the project. The use of Patton's *Utilization-focused Evaluation* (UfE) complements and strengthen EE.

The focus in UfE is brought directly to the level of the internal and external stakeholders who will benefit from a programme. Developed by Michael Quinn Patton in the 1980s, *Utilization-focused Evaluation* (UfE) systematically draws attention to the users of programme evaluations by internal *and* external stakeholders. UfE strives to firstly identify the people who will use the information garnered in the evaluation and then actively involve them in making decisions with regard to its roll-out. It argues that evaluation should be judged by its utility and actual use. The process should take careful consideration of how everything that will be done, from beginning to end, will affect use. In considering this, evaluators' consideration of the context and more importantly the culture within which the evaluation is taking place must be central. UfE would urge evaluators to view culture from a strengths perspective and through this, establish meaningful discourse about shared understandings and robust training in evaluation, utility, feasibility, propriety, accuracy (Patton, 1986). Through this shared understanding, Kirkhart (2010) argues that there is the potential for cultural congruence in this model of evaluation, as it operates in a very rich contextual way. UfE is essentially a situational analysis and focuses on both *what works* and *what makes sense* in that particular situation. Its strengths, she argues, are its emphases on fluidity and complexity. There is '*no one road to cultural congruence, but it can be established through the application of theories that are culturally specific or broadly define by overarching purpose'* (Kirkhart, 2010, 410). Its use, according to Patton et al, does not occur naturally or automatically; the groundwork for it has to be carefully laid. UfE, as a fundamental orientation, applies a five-step framework to the evaluation process:

1. Identify primary intended users

2. Gain commitment of key stakeholders

3. Decide on evaluation options

4. Analyze and interpret findings, reach conclusions

5. Disseminate findings.

(Patton, 1986, 2001, 2002, 2008)

This evaluation approach puts a strong emphasis on the key stakeholders (Bryson, Patton & Bowman, 2011).   UfE insists that the actual use of an evaluation is not necessarily the same as elements that are common to most evaluation processes, i.e., production of reports and dissemination. As the stakeholders are involved in the design of the evaluation, they will know best how to disseminate it and to whom.

Both perspectives discussed above highly value the users' perspective. EE is guided by principles of self-determination and capacity building. It places the user/stakeholder in charge of their own decisions. It is at this overlap of the two perspectives that the current project resides (Figure 1).   It is a challenge to achieve cultural congruence within a highly diverse, multi-country project, but a strong situational analysis of each country's stakeholders allows for fluidity and allows the complexity of the programme to be evaluated. A situational analysis was facilitated by thematically analysing data from each country in relation to the local needs of that country and how impact was defined by the partners in that country. The user of the project was at the centre of this analysis. This is alluded to in the section which discusses the procedure.
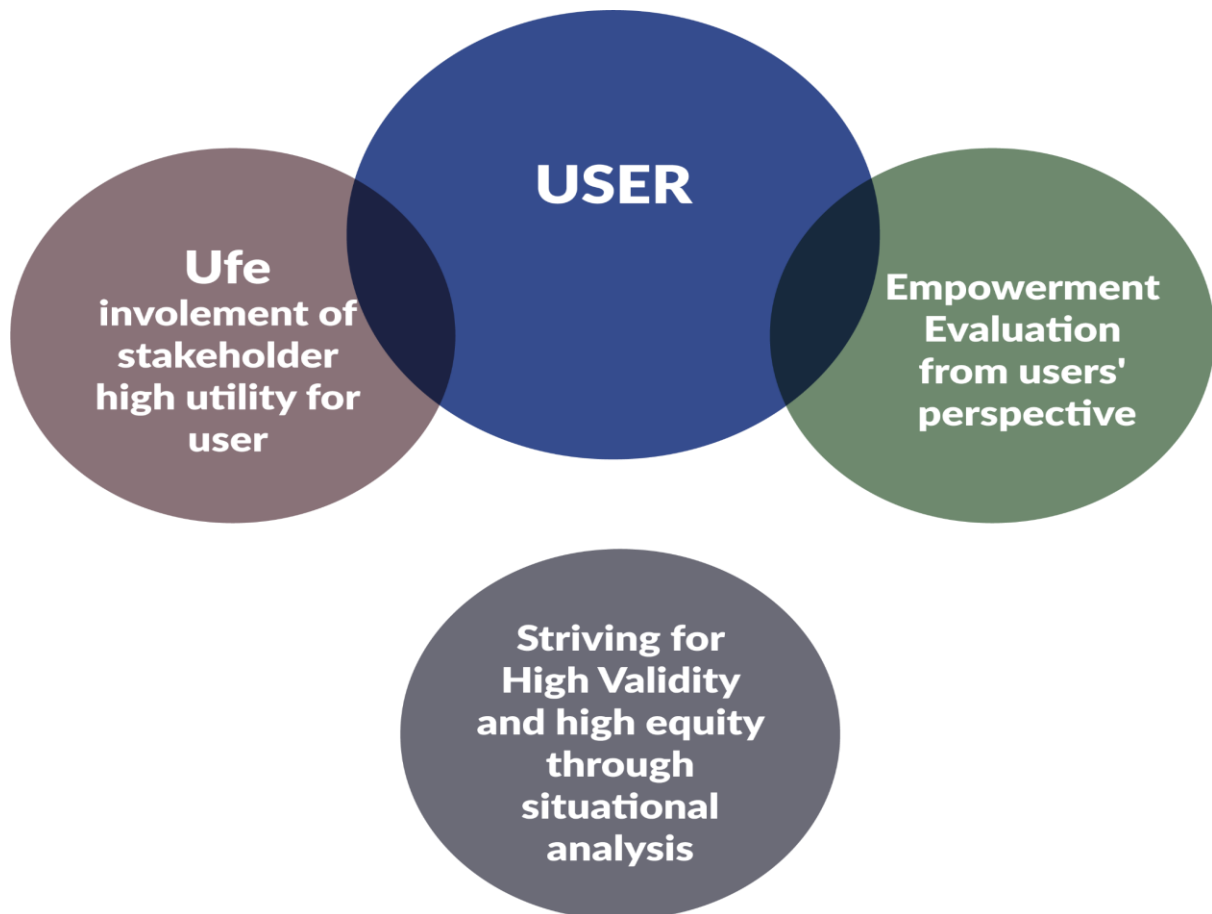

Figure 1 Conceptualisation of Evaluation in current study

Empowerment Evaluation and *Utilization-focused Evaluation* both draw on theories of action and organisational learning that emphasise the necessity – and capacity – of organisations to adapt to critical developments in their context. Most prominently, this has been expressed by Donald Schön and Chris Argyris in the concept of *Double-Loop learning* (Argyris & Schön, 1978, 1996). Most organisations regularly engage in what Schön and Argyris call *Single-Loop Learning*: they ask what they need to do in order to better achieve their stated goals. While useful to some extent, such activities often take the organisation's goals or mission for granted, therefore carrying the risk of ignoring crucial developments and changes in the organisation's wider context and environment (think Nokia's response to Apple's invention of the smart phone). *Double-Loop Learning* involves a process of asking critical questions about the taken-for-granted assumptions the organisation or programme is based upon, enabling re-adjustment and re-invention.

The evidence is reasonably limited in terms of EE and UfE's suitability among certain contexts. The current study explores empirical findings around the potential of both approaches to provide working principles in a cross-cultural project.

The current study

The research team was charged with conducting the impact evaluation of *TOY to Share, Play to Care* (www.toy4inclusion.eu/). The objective of this second phase of *TOY for Inclusion* was to broaden the target beneficiaries, scale up and embed in policy the *TOY for Inclusion* approach. This has led to the establishment of sixteen community-based early childhood education and care (ECEC) play hubs in eight European countries, reaching more than 10,000 children, 5,000 adults and 1,000 practitioners. These play hubs are spaces where young children from minority ethnic and marginalised communities can come to play with their parents and grandparents, with many of the hubs providing a toy library service also. The play hubs are located within the community so to be accessible for all families. Each play hub is supported by the national *TOY for Inclusion* partner organisation but is managed locally by a multi-sectoral Local Action Team (LAT) comprising of a project co-ordinator and community representatives (i.e. school and preschool teachers, health, social and early childhood education and care service providers, parents and local authority personnel).

TOY to Share, Play to Care builds on the expertise and experiences gathered in the first phase of *TOY for Inclusion* and uses such findings to inform and guide the establishment of new play hubs and expand the outreach of the existing play hubs over the 20-month time span of the project. To achieve this, at a first level, learning *from practice for practice* is systematically built into the project design. At a second level, *TOY to Share, Play to Care* encourages reflective learning *about practice*, with experienced and new LAT co-ordinators engaged in critical self-reflection of *TOY for Inclusion* activities. Finally, at a third level, the project, through its documented evaluation, enables *policy learning* beyond the immediate *TOY* context, thus offering pathways for potential systemic impact on the future development of integrated early childhood education and care programmes.

The research team designed and conducted the impact evaluation as a work package among several other work packages which informed the design and development of *TOY to Share, Play to Care*. From January 2019 to August 2020, the impact evaluation focused on a number of participatory activities to collect the data and consultation with the other work packages (outlined in the section on Procedure and Findings). Each activity of the impact evaluation procedure was designed keeping in mind a) the overall progress of the project and b) the goal of providing policy recommendations at

various levels (local, regional, national, EU) as a key outcome of this cross-cultural project. Considering the complexity of the project, with a multitude of actors and stakeholders in markedly different locations in Europe, the nature of the impact evaluation was necessarily processual, allowing for adaptation and adjustment as the project evolved. The challenge for the research team was for us to focus on the overall objective of *TOY to Share, Play to Care,* while negotiating a balance with the inevitable unpredictability and uncertainty that characterises all learning in and from evolving and complex situations. We consider *unpredictability* to be a potentially productive force of a project like *TOY to Share, Play to Care*. Therefore, we oriented the design of the impact evaluation towards the *desired impact* as stated in the project description as well as the *unintended* and *surprising* experiences that we expected to occur in order to make them accessible for more generalised policy and practice learning. The impact evaluation was designed and carried out as one of three elements of the overall evaluation framework of the project:

- Initial and continuous documentation and monitoring of local capacities and needs (*Stakeholder Mapping*), Play Hub usage, reach and activities (*Data collection* and *Monitoring* protocols and tools)

- Empowerment and Utilization-focused *Impact evaluation*, documenting participants' experiences and their views on what supports or constrains *making a difference*

- External evaluation of project conduct and achievement

The strong contextual nature of the project meant that it needed to utilise a suitably collaborative approach in its impact evaluation with the 'user' at the core of the process. UfE and EE enabled this to happen and we outline here the integration of both approaches and the innovative aspects of our impact evaluation to inform future evaluations and to share our learning.

The Impact Evaluation: Procedure and Findings

The 'users'[1] were identified as central to this evaluation. The users in the case of this impact evaluation were identified as high-level partners and stakeholders with responsibility for decision-making in each country. The user group consisted of the national *TOY for Inclusion* partner organisations in the eight participating countries, the local action team co-ordinators and any members of that local action team. The aim of the recruitment at this stage, was primarily to get as many representative voices at the table as possible. In keeping with the conceptual framework and overall research design of the Work Package–*Empowerment Evaluation* and *Utilization Focused Evaluation*–the type of data that was collected was judged by its potential utility and actual use in the project. To facilitate *Double Loop Learning,* timely feedback was given to partners engaging in the stakeholder mapping process and partners with responsibility for the development of the templates for monitoring data in the project. In addition, preliminary findings were presented in an interim report that was shared with all partners for comments and feedback.

---

[1] In this paper we refer to 'users', 'participants', and 'stakeholders'. The terms intersect but are not identical. For the purpose of this paper 'user' is the overarching term with reference to the conceptual framework of the evaluation. 'Participants' are those actively involved in the impact evaluation activities while 'stakeholders' refers to a wider group of individuals and agencies with an interest in the project, its conduct, and its outcomes (e.g. policy makers).

Building on the initial proposal for the impact evaluation, and following introductory discussions with project coordinators and project partners at the kick-off meeting in Rome in March 2019, our approach was to centre the evaluation on the perceptions of local stakeholders. In consequence, the general term *impact* is framed in this evaluation as *making a difference in the locality* in relation to the overall aims of *Toy to Share, Play to Care*. In keeping with this overall focus, we organised the collection, analysis and interpretation of data according to three main research questions:

1. What does impact/making a difference mean to your locality in relation to inclusive early years community initiatives i.e. this project?

    a. How do you know?

    b. For whom?

2. What do you envisage will help you make a difference to your locality in relation to inclusive early years community initiatives?

3. What do you envisage will make it difficult to make a difference in relation to inclusive early years community initiatives?

Procedure

Ethical clearance was obtained from the Institution's Research Ethics Committee for this impact evaluation. Plain language statements were provided to all participants and informed consent was obtained.

Stakeholders

EE's three steps of: *establishing the mission statement, taking stock and agreement of future goals* informed the data collection plan (Patton, 2001, 2008). Participants included:

● International *TOY for Inclusion* Project Managers

● National *TOY for Inclusion* partner organisations from eight participating countries: Hungary, The Netherlands, Slovakia, Italy, Turkey, Croatia, Slovenia and Latvia

● Experienced Local Action Team (LAT) Co-ordinators from *TOY for Inclusion*

● New Local Action Team (LAT) Co-ordinators

● Romani Community Cultural Mediators

● *TOY for Inclusion* Project Trainer

● External evaluator for *TOY to Share, Play to Care*

● Communications officer on *TOY to Share, Play to Care*

Data Collection

Establishing the Mission Statement

The first stage of data collection took place in May 2019 at a *Training of Trainers* event in Croatia. During this three-day event, the impact evaluation team undertook observations of the training

events, conducted semi-structured interviews, focus groups, and task-based interviews with participants, i.e. users and stakeholders.

As this project was built on the initial phase of *TOY for Inclusion*, the mission statement was already created and in use. This mission statement was repeatedly and clearly enunciated to partners in the project at various training sessions and partner meetings, as witnessed by the external impact evaluators, during data collection observations. However, in order to allow the stakeholders to scrutinise it, the mission statement was interrogated by the key partners during the interviews and focus groups (Figure 2).
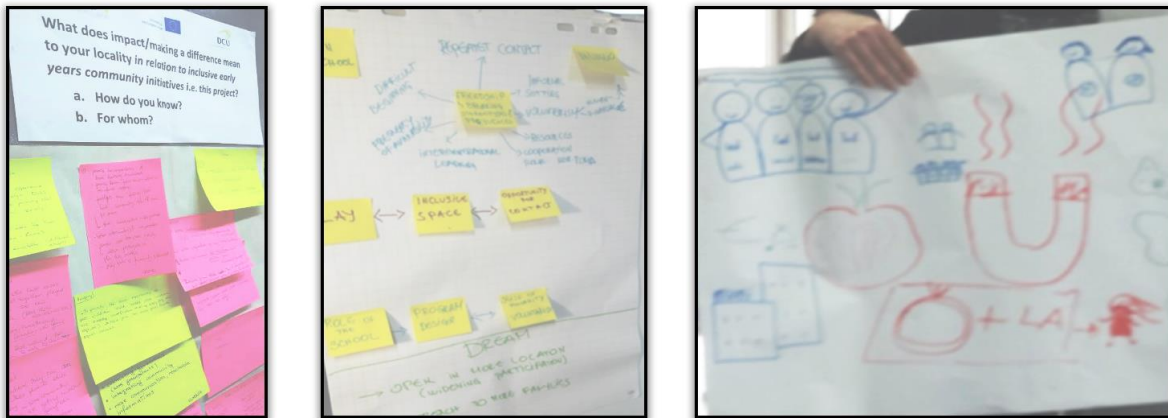
Figure 2: Examples of the data collection, including documented focus group discussions and task-based activities at the Training of Trainers event in Sisak, Croatia, in May 2019.

For both face-to-face and online engagements with participants, the following data handling protocol was adhered to:

- To increase reliability of the interview data, interviews were audio-recorded and conducted by the same researcher with a note-taker present. Recordings were transcribed and examined and verified against written notes in order to ensure validity of participants' responses. This was particularly important, due to English being the medium through which the interviews were conducted, and many of the participants having English as an Additional Language.

- The same rigour was applied to the focus group data. Note-takers were present in all focus groups and the data was examined immediately after the focus groups, to assure accuracy of representation of the responses.

- Task-based interviews and activities were conducted to examine the stakeholder mapping, peer mentoring and monitoring paperwork. This data was organised immediately after the activities, in order to increase its reliability.

- Observation data was collected by all researchers through field notes during the training event.

- All data was entered into *QDA miner* qualitative data analysis software to facilitate coding and in-depth analysis across varied data sources.

Table 1 Summary of Data Collection

	Experienced LAT co-ordinators	Partners & Others (Romani cultural mediators, trainer etc.)	New LAT co-ordinators	Communication Officer Project Manager External Evaluator	All participants at Training event
Semi-structured Focus Group	√	√	√		
Task-Based Interviews (Stakeholders, Peer Mentoring & Monitoring Data)	√	√	√		
Observation Data					√
Semi-structured Interviews			√	√	

The evaluation very quickly moved to step two: *taking stock*.

Taking Stock

At the halfway point of the project (Stage 2), participants evaluated their participation through interviews and examined their own play hub's monitoring data. Mixed focus group participants were asked to reflect and examine their own participation in the project. In addition to this, further interviews were originally scheduled to take place in March 2020, at a project meeting in Slovakia. However, due to the early stages of the Covid-19 pandemic the face-to face meeting had to be cancelled and moved to online. As a result of this change, interviews with coordinators of the newly established Local Action Teams were conducted online, using Zoom videoconferencing software.

Taking stock was a cyclical process, with double loop learning engaged. This strengths-based approach was centred on dialogue. The evaluator and the participants were equals in the process, operating as a critical friend, as opposed to a party that offers critique only. The decision around documentation or what evidence is produced was one that was taken as a group. Participants involved in two other work packages (those responsible for monitoring data and those responsible for stakeholder mapping) were offered feedback on the usefulness of these documents by the users themselves along with the impact evaluation team. Through this double loop learning shared understanding was developed and ownership of the monitoring and stakeholder evaluation process was shared amongst the actors. All participants were also invited to draw on local artefacts and pieces of evidence from their local area and use them to inform the discussions with the evaluators around the utility of the programme. It was hoped that by engaging in real-context examples, deeper feedback could be discussed, which would result in optimum sustainability of the project.

Agreeing Goals for Moving forward

The third step: *agreeing goals for moving forward* was facilitated through focus group discussions and semi-structured online interviews as outlined above. In addition to this, a narrative report was produced by the evaluation team at the halfway point. Partners and project leads were

instrumental in this report and feedback was incorporated into ongoing planning. Narrative report produced at the end of the project also incorporated all partners' feedback, in the hope that this could be used in any future iterations of the project.

Data Analysis and Interpretation

The data analysis in this impact evaluation is qualitative and interpretative by design, to ensure the appropriate level of depth required for a complex context like *Toy to Share, Play to Care* (Merriam, 1998; Yin, 2006). The research analysis was guided by the conceptual framework of *Empowerment Evaluation* and *Utilization-focused evaluation*, and an inductive process was employed whereby the analysis was driven by the data itself. *Empowerment evaluation* provides participants with opportunities to self-evaluate (Fetterman et al, 1996) and *Utilization-focused Evaluation* considers how the process of a project from being to end will affect use (Patton, 1986).

A qualitative data analysis software (CAQDAS) package, (QDA Miner) was used to assist in the thematic data analysis process. CAQDAS software facilitates data storage, retrieval, coding, comparison and making connections (Patton, 2002). Such software packages enable researchers to develop an accurate and transparent picture of the data whilst also providing an audit of the data analysis process as a whole. Vigilant systems of data collection are required to enable rigorous analysis.

QDA Miner software was used, therefore, to facilitate re-reading, sorting and retrieval of codes and themes to ensure high levels of inter-reliability among the research team. Coding is one of the significant steps taken during thematic analysis to organize and make sense of textual data. Thematic analysis is a method for identifying and analysing patterns in qualitative data. The analysis was completed, drawing from the six phases of thematic data analysis outlined by Braun and Clarke (2012):

1. ***Familiarisation of the data***: The research team immersed themselves in the data emerging from the semi-structured focus-groups, post-it data, the workshop data, the task-based interviews and the observation data (day-to-day) initially. The interview recordings were re-listened to and transcriptions were read and reread.

2. ***Generating initial codes:*** Initial codes of the interview data were developed.

3. ***Searching for themes:*** The researchers then engaged in a process of '*reflecting upon their actions and values during research...and the effects that they might have'* (Robson, 2002, p. 551). The team generated proposed themes and a corresponding codebook was developed.

4. ***Reviewing potential themes:*** The team then worked in pairs and the codes were then applied across all the data (post-it data, semi-structured focus-groups, workshop data, task-based interviews, observation data and semi-structured interviews).

5. ***Defining and naming themes:*** The dataset was further analysed and the codebook further refined.

6. ***Producing a preliminary report:*** Finally, coded and categorised text was exported to Microsoft Excel to allow for coding frequency and selection of illustrative examples. See Figure 3 below for a table demonstrating coding frequency and its variation from Stage 1 to Stage 2
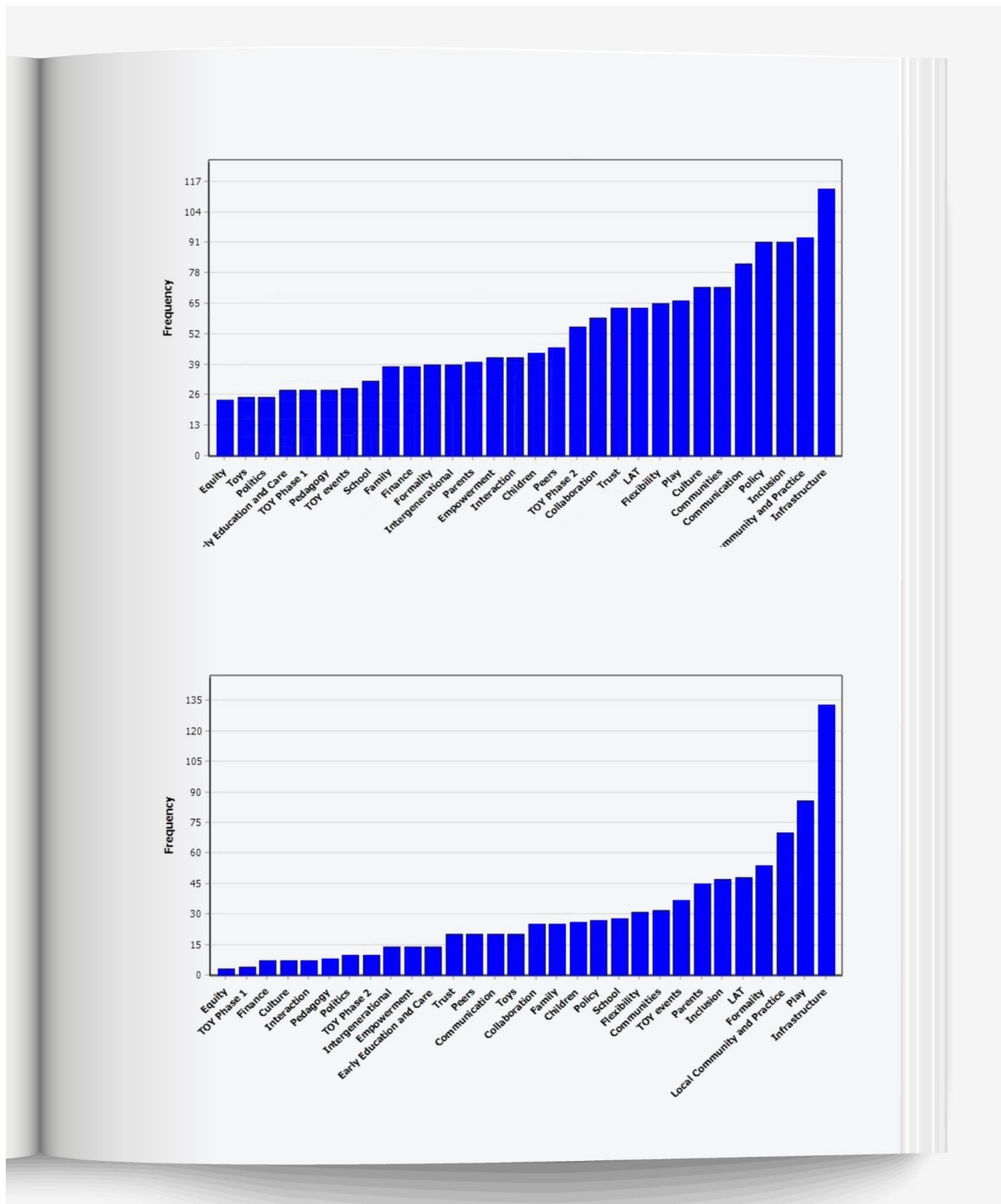
Figure 3 Coding frequency and its variation from Stage 1 to Stage 2

To illustrate the initial thematic analysis, the most frequent theme in both stages of the data collection is 'Infrastructure'. The theme of Infrastructure refers to the following initial codes: *Space/location, resources, human resources, facilitation by municipalities and policy makers, transport, access to Play Hubs/communities, routines and 'rules' of Play Hub, operation of Play Hubs*. Beginning with the initial analysis of data collected in Stage 1, this coding was applied across all data sets and identified 114 separate mentions, emerging from 5 data sets. The following are examples of what emerged from those 114 responses during the initial analysis:

Table 2: Example of initial data analysis

Data set	Response
Post – it Data	"People with energy and will."
Semi-structured interview	"Our play hub is located in an integrational board department."
Day to Day 1	"LAT team meets once a month. All skilled and all professionals."
Focus-group	"The location is important."
Workshop data	"Cooking pot: Bowl represents LAT team/Play Hub with ingredients such as members, school, teachers, parents, Roma, health sector, social, volunteers."

Following the initial analysis of data collected in stage 1 (May 2019) and stage 2 (March 2020) we moved from first-level analysis (e.g. frequency of mentions) to second-level analysis. This involved re-visiting the initial coding, and moving to a deeper analytical interpretation of *meaning* and underlying discourses. This enabled us to elicit is how the thematic nodes connect to the main research questions guiding the impact evaluation (see above). This allowed us, for instance, to show the relationship (and possible difference) between values held by the *TOY* consortium and partners (e.g. children's rights, participation, inclusion) and hands-on priorities arising from the task to set up and/or scale up a community-based early childhood programme under often difficult conditions.

Following transcription of all material (audio-recordings, observation notes, 'post-it notes' generated in task-based activities) and initial coding (see above), the material was condensed into 109 pages for second-level coding and detailed analysis. Figure 4 below gives an example of the template developed and used for this phase of analysis.

Figure 4: Template for detailed thematic analysis

Taking this approach enabled us to clearly link the data, via the two levels of thematic analysis, to the initial research questions. The findings of the analysis are summarised in the next section.

Impact Evaluation Findings

As outlined above, we organised the data and analysis in a way that facilitated the two models of evaluation discussed above, as well as taking into account some opportunities for double loop learning.

The advantages of taking this approach to the impact evaluation of this project were twofold. First, it allowed for critical reflection *within* the *TOY to Share, Play to Care* community, to revisit their own processes and make use of it for future adaptations both at local and overarching project level. The users learned about their own values system, about what impact means to their local context, and by doing this, made changes to the project during and for future iterations of the programme. This was a situational approach and brought the analysis to the level of the user. Double-loop learning was a crucial part of this, with feedback provided on the stakeholder mapping and the monitoring templates that each partner used for their play hub.

Second, linking the analysis to the three clearly defined questions enables interpretation of the findings for *external* use, e.g. to inform future policy choices at local, national and European Union levels.

Stakeholder perspectives

The first research question addressed stakeholders' own views and perceptions on what *making a difference* is all about. Responses to the question frame topics that dominated the conversations in anticipation of the setting up of new Play Hubs (new participants) as well as the reflection on previous experiences (participants from existing Play Hubs). The stakeholders identified both local and global factors that define impact for them, thus engaging in a situational analysis of their own context. Identifying the common thematic nodes across users, enabled us to generate common ground for both internal and external use in the programme. This resulted in the users of the programme deeply reflecting on their role in the programme and how that interacts with both the mission statement and overall aims of the project. Empowerment was taking place at the individual level in the appraisal of individual roles in the programme.

Supportive factors

The second research question focused the users on supporting factors that contribute to the success of the *Play Hubs*. Reading the data, we identified themes (thematic nodes). Stakeholders reported that relationships and local engagement were central to the development of trust in their local contexts. They identified that the use of personnel from within the user group was a key supportive factor to its successful roll-out. The stakeholders were best placed to identify the supportive factors that could not only increase engagement from their local communities but also provide sustainability of it into the future.

Barriers to making a difference

Research question three aimed at gathering information about factors that impede and hinder local impact, according to the stakeholders themselves. As with research questions one and two, the

topics (thematic nodes) emerging under this heading are based on participants' perceptions and own experiences. Having stakeholders examining their contexts for what might impinge on the project's success helped to create pragmatic modes of addressing these barriers and avoidance of them entirely going forward.

Double Loop learning - usefulness of the monitoring framework

Double loop learning by its nature denotes that the stakeholders engage in critical self-reflection. This enables them to feed forward any observations of the programme and how it is working on the ground while it happens, rather than waiting for the end of the project. It increases the chances of success of the project, as it highlights issues and address them in pragmatic ways offered by the stakeholders themselves. Stakeholders gain the opportunity to try and test these potential solutions as they arise. In this project, monitoring templates were examined by the project partners and local action teams.

In line with the aim and conceptual orientation of this *impact evaluation* (Utilization-focused Evaluation), along with the designers and users, we analysed the data for perspectives on the *use* and *usefulness* of these instruments. The *impact evaluation* was interested in documenting whether (or not), and how, the instruments provided by other work packages support project participants in their main goal of *making a difference* and, with a view to extracting possible policy lessons from the project, how their experiences can be put to use for future projects with similar aims.

In general, participants expressed positive views on the monitoring frameworks and the templates. There was an overall agreement that it was useful to have a shared framework to document activities of the local hubs, and the templates that were provided for recording attendance and other usage data were welcomed. Repeatedly, participants expressed that collecting and reviewing the monitoring data had been 'revealing', and 'foundational in the development of services'.

However, some critical aspects emerged over the course of the project, as well as suggestions for change and improvement. Participants reflected on the monitoring framework both from a local perspective (usefulness for the development and operation of own *Play Hubs*) and with reference to the overall *TOY to Share, Play to Care* project.

Reported experiences with the monitoring framework

Positive statements about the monitoring framework tended to centre around planning, organising, communicating and reviewing on one's own practices and on reporting local developments to the project coordination and contributing to the evidence base of the overall project.

Despite the overall positive perspective on the monitoring framework some critical views emerged. They are mainly related to two aspects: 1. Time and workload demands created by both the collection and the analysis of data, and 2. Some more fundamental questions about how (and by whom) the data would be used.

Based on their experience with the monitoring framework and tools participants suggested (or implemented) several amendments. They were generally pragmatic changes to the way data was recorded with the help of the initial templates provided by the project. All of these insights provided really important feedback data to the work packages that had responsibility for creating and maintaining the data templates. The data was used to empower the stakeholders towards informed

decision-making while the project was taking place. Alongside that, the external evaluation team, us, worked on providing our own feedback, which was also helpful in creating better and more useful templates for the hubs

Discussion

This impact evaluation set out to establish what *making a difference* meant to the local participants in a diverse, multi-country context. This meant the evaluation of the impact of a programme in complex real-world contexts. We employed two major models of evaluation of *Empowerment Evaluation* and *Utilization Focused Evaluation* to underpin this and found a common ground at the point of the user. This ensured that the user or stakeholder was at the centre of the impact evaluation and the feedback and findings could be useful, even during the project, as per the model presented earlier in Figure 1. This supported the empowerment of the user.   In striving for high validity and equity, there were challenges along the way. These can be listed by way of local challenges to the project with regard to evaluation and, in turn, global and structural ones.

Local challenges

Fetterman's three principles of Empowerment Evaluation (EE) were broadly followed;

- creating the mission statement,

- taking stock

- and forward planning.

However, addressing the first of these principles was a challenge, as is often the case in evaluation of real-world large-scale projects. The project was the upscaling of a previous project and had already commenced by the time the evaluation team were in place. As a result, the mission statement could not be developed alongside the evaluation team or all stakeholders. However, the mission statement was carefully scrutinised by way of the main research question: *What does making a difference mean to your local area?* Some stakeholders were new to the project, so this enabled a fresh look at the mission statement and an onus on existing stakeholders to explain and justify its suitability to such a project. This self-reflection on the mission statement resulted in useful discussion among the partners on potential problem-solving techniques and problem-avoiding in their own countries.

The underlying assumption for this approach is that *impact* and *agency* are neither neutral nor universal categories. On the contrary, they are always and inevitably tied to the situations in which participants exert their agency. *TOY to Share, Play to Care* is a complex project. As the context for each local hub is different and unique (in the overall frame of the project), so are participants' perspectives on what matters, what counts, helps and hinders for the local area. Each area's needs are localised. But the approach employed in this evaluation capitalised on this local challenges and series of challenges to allow the users to scrutinise what the best course of action for the project for their particular area is. This could have resulted in a proneness to bias if left unchecked (Scriven 2017). This was addressed in this impact evaluation by the team of five external evaluators who could examine the data, objectively analyse it, while still offering it back to the stakeholders for validation.

Structural or global barriers to EE and UfE

Alongside the local challenges, a main barrier to a deep dive into the impact of the project was language. Although the common language of the project was English, many of the LAT or local workers to the project did not speak English or had English as an Additional Language. This was difficult then in terms of engaging in evaluation exercises, as the aims and overall mission statement would have been originally drafted in English. The nuances of language are such that a word or term (e.g., early education) could mean something in one language, but then look very different in another context. Often the terms were being translated and explained by interpreters. Also, the reflection questions being posed by the evaluators were often being translated. It is possible that the nuances of the respondents' answers could have been lost through the translation. However, the stakeholders in the current impact evaluation had the opportunity to view drafts of the impact report as it was being written. They had the opportunities to feed back about the findings and most importantly ask the external evaluators questions about the nature of the data reported. This resulted in a more balanced impact report with input from a wide range of stakeholders.

A second global challenge was, of course, a global pandemic. This inhibited, in some cases, the full roll out of the programme itself. This meant that in 'taking stock' during the interviews, some of the participants had not fully had the opportunity to try everything out, so could not fully offer feedback or reflection on their participation in particular aspects of the programme. The groups could not get together, i.e. external facilitators and internal programme users. These unforeseen events can inhibit the evaluation of the programme in its purest sense, but it can also offer opportunities to self-evaluate in other directions. Nobody could have foreseen this, but how the users responded to such a crisis could be evaluated. How they operated the Play Hubs in such a crisis and how the project functioned with such an unexpected set of circumstances was very important data for both the programme team and the facilitators at a local level. Reflecting on the action taken, can prove insightful and useful. These unexpected consequences of the pandemic actually provided a truer sense in one way of UfE, as the evaluation had to shift and change with the new circumstances it was in. This was valuable to the stakeholders which then fed directly into the future goals for moving forward. New information about mitigating circumstances could be fed forward and held for maintenance of the programme beyond the challenges of a global pandemic. The challenges could be locally based but more general principles could be developed based on this.

How the models worked successfully

The models worked in tandem at a meta-level with the user at the centre. They provided structure to what could have been a challenging impact evaluation due to the very different local contexts that existed in the project. This approach provided an opportunity, for a road to cultural congruence and an identification of some systemic factors that could be applied and useful to the programme going forward.

For example, the systemic factors that emerged can be broken down as outlined in Figure 5. These emerged directly from the stakeholders themselves and included what is denoted in the figure below.
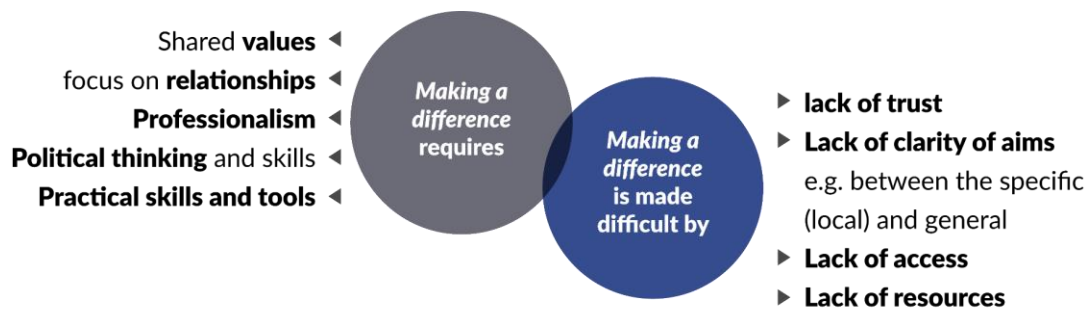
Figure 5 Systemic factors

The participants identified shared values as being incredibly important to the success of any programme in their local context. These values emerge from the ability to develop strong relationships built on trust with the local community. As the project partners were working with quite marginalised communities, trust was key to the communities getting involved at all and actually coming to the play hubs. The participants knew this and made great leaps and bounds in order to get them there. They moved the play hubs to the communities themselves and even offered a mobile play hub in one jurisdiction. This continued to be a key factor when the pandemic hit also. The leaders had to essentially think on their feet in order to pivot, in some case, to an online provision.

The professionalism of the providers was also identified as part of the values that supported a successful programme as they define it. The emphasised the importance of having professionals involved, but also utilising the professionals that exist within the communities themselves and the skillsets that exist within the communities, for the communities. They spoke about the importance of political will and political skill in creating a project that will sustain and maintain the effects or impact it is having on the local community. They said that without the support (financial or otherwise,) the project could not be sustained.

What is even more important, is that they identified what will work for them in their local context. This is where they were truly being empowered to make change and use the information to sustain the project. They knew the pitfalls from their discussion and reflection which they had to avoid in order to have the project fail. Things such as a lack of trust among stakeholders can create uneasiness and does not facilitate the development of relationship, which is key if communities are to actually come to the service. Secondly, a lack of clarity or aims for the overall project, but also for each individual hub is important. But between those two contexts also. If the aims are not communicated and unifying, then the project will be steered in one way other than what might be useful to the community. So the identification of stakeholders' needs from the beginning is incredibly important.

 Of course, a lack of resources can be seen as a major barrier. By the stakeholders identifying what can give them success, they can ensure that these things are in place before any further iterations go ahead. So ensuring that the resources such as buildings and infrastructure and links to already established community holdings such as schools was particularly important for this project.

Brought together in one picture the above dimensions and elements point to a number of systemic factors that either support or hinder local impact. Both sides are present in the *TOY to Share, Play to*

Care project (as expressed by participants in the evaluation); they should and can be kept in mind for both future directions the project may take, and for the policy lessons that can be learned beyond this specific project. They include both *soft* (relationships, trust, clarity) and *hard* (resources, access, skills) aspects, with *professionalism* appearing to be key in successfully negotiating the factors.

It is becoming increasingly clear that in order to be successful, projects that combine local diversity in a shared framework of values and principles require a specific approach to critical reflection across all elements of the project. In an evaluation of early childhood development programmes carried out for the Bernard van Leer Foundation ('The effectiveness initiative') Leonardo Yanez stated that the most effective programmes are those that have space, time and resources for critical reflection built into their approach (Bernard van Leer Foundation, 2001). While reflection has long been acknowledged as a critical factor for the success of projects, *TOY to Share, Play to Care* indicates the need for an advanced understanding of the role of critical reflection. While necessary, it is not enough to require project participants to reflect on their own, local practices (or evaluate the 'implementation' of centrally devised programmes at local level). Instead, the project has to introduce measures that ensure learning from local experience is systematically fed back and used to critically question, reframe and recalibrate the assumptions that orient the overall project. Argyris and Schön (1978) refer to this as *double-loop learning*. Another way of putting it is to shift the focus of reflection from the question *are we doing things right?* to *Are we doing the right things?*

Both the challenge *and* the possibility for a project like *TOY to Share, Play to Care* arises from this necessary shift from approaches that seek to *implement* programmes to approaches that actively encourage local *interpretation* and situated *meaning-making.*

To systematically learn from these processes will support the development of highly effective *grassroots competent systems* (Author et al., 2020b; Author, 2018)

Beyond the project context, some obvious challenges for policy makers arise from such systemic re-orientation of approaches. They include a radical rethinking of *governance* as distributed, something that is intrinsic to the system and all its actors, rather than top-down regulation, steering or control. *Implementation* (of policies and programmes), in scenarios of distributed intrinsic governance, transforms into *interpretation*–i.e. actors exerting judgement and *making sense* of policies and programmes based on their own expertise. This, in consequence, requires approaches to governance and policy implementation that are grounded in *trust* in the creative power, competence, and professionalism of all. Central to achieving this transformation is the systematic introduction–and resourcing–of critically reflective cycles (*double-loop learning*) at all levels of the system including the sphere of policy making.

Lessons Learned

Findings from the current project show that the models of EE and UfE can be adapted and used in a real-world context. By putting the user at the centre, projects can plan for future sustainability with the surety that what they are doing is useful and making a difference to the users in local contexts. This type of impact evaluation allows central forces to listen and build relationships with the local stakeholders. This project has shown that through these mechanisms it is possible to develop sets of

principles that are useful to policy makers at national and international level or a range of unexpected principles. In keeping with Patton's five working principles of UfE:

1. Identify primary intended users

2. Gain commitment of key stakeholders

3. Decide on evaluation options

4. Analyse and interpret findings, reach conclusions

5. Disseminate findings.

This project followed these with some reasonable success. The users were identified through a local stakeholder mapping. The impact evaluation fed back to this process and this helped to identify current and potential users of the project. Relationships between the external and internal evaluators were built and sustained, through attendance at both training and partner events. Commitment of the stakeholders to the evaluation and the evaluation team was gained. The evaluation options were decided upon with the project leads. Suggestions were made by project partners as to how the evaluation might be conducted. If practicable and with agreement, these were incorporated into the overall evaluation design. The findings were very robustly and rigorously analysed with an emphasis on inclusion and data verification by the participants.

Final Conclusion

Conclusions were reached and shared with the users and feedback on these findings was considered carefully by the external evaluators. This feedback was incorporated as far as possible, while not interfering with the integrity of the independence of a rigorous impact evaluation. The evaluation in this study adhered more closely to Patton's model of UfE than EE. EE was attempted earnestly, but was constrained by the fact that the project had had a very clear mission statement already in place. The users were not therefore in a position to have input to this. However, stakeholders did benefit from critical reflection on it in relation to their own local contexts.   This is where the main conclusion can be drawn from this work. Local level reflection has allowed for more general and far reaching principles to be drawn up. These can be used in a more wide-ranging way across further projects that address complex local needs but on a Europe wide and broader international policy basis. The findings and principles resonate with what Author et al (2018) describe as an emerging 'systemic turn': recent policy developments at international level that respond to increasing complexity and ambiguity in the governance of early childhood systems, especially under crisis conditions (Council of the European Union, 2019; Author et al., 2020a; Author, 2018; Author, 2012).

Acknowledgements

References

Argyris, C., & Schön, D. A. (1978). *Organizational learning*. Reading, Mass.: Addison-Wesley Pub. Co.

Argyris, C. & Schön, D.A. (1996). *Organizational learning 11: theory, method, and practice*. Reading, MA: Addison-Wesley.

Bernard van Leer Foundation. (2001). *The effectiveness initiative: First fruits*. The Hague: Bernard van Leer Foundation.

Author,. (2012). Towards competent systems in early childhood education and care. Implications for policy and practice. *European Journal of Education, 47*(4), 508-526. doi:10.1111/ejed.12010rogplan.2016.10.004

Author, (2018). It takes more than a village. Effective early childhood development, education and care services require competent systems. *Global Solutions Journal, 1*(2), 116-124.

Author, (2020a). *Post-Covid-19 to 2030: Early childhood programs as pathway to sustainability in times of global uncertainty*. Retrieved from Riyadh, Saudi Arabia: https://t20saudiarabia.org.sa/en/briefs/Pages/Policy-Brief.aspx?pb=TF7_PB3

Author, (2020b). *Upscaling community based early childhood programmes to counter inequality and foster social cohesion during global uncertainty*. Retrieved from Riyadh, Saudi Arabia: https://t20saudiarabia.org.sa/en/briefs/Pages/Policy-Brief.aspx?pb=TF4_PB5

Author. (2021). *TOY for inclusion: Impact evaluation and policy recommendations. Final report*. Dublin: Author research centre.

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA handbooks in psychology. APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological (p. 57–71). American Psychological Association. doi:10.1037/13620-004

Bryson, J.M., Patton, M.Q. & Bowman, R.A. (2011). Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, 34, 1-12.

Council of the European Union. (2019). *Council recommendation of 22 May 2019 on high-quality early childhood education and care systems*. (2019/C 189/02). Brussels: Council of the European Union Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2019.189.01.0004.01.ENG&toc=OJ:C:2019:189:TOC

Cousins, J.B., Whitmore, E. & Shulha, L. (2012). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation*, 34(1), 7-22. doi: 10.1177/1098214012464037

Fetterman, D.M., Kaftarian, S., & Wandersman, A. (Eds.). (1996). *Empowerment evaluation: knowledge and tools for self-assessment and accountability*. Thousand Oaks: Sage.

Fetterman, D.M. (2001). *Foundations of empowerment evaluation*. Thousand Oaks: Sage.

Fetterman, D.M. and Wandersman, A. (2005). *Empowerment evaluation principles in practice*. New York: Guilford Publications.

Fetterman, D.M. and Wandersman, A. (2007). Empowerment Evaluation: Yesterday, Today and Tomorrow. *American Journal of Evaluation,* 28(2), 179-198.

Fetterman, D.M. (2013). *Empowerment evaluation in the digital villages: Hewlett-Packard's $15 million race toward social justice*. Stanford: Stanford University Press.

Fetterman, D.M., Kaftarian, S., and Wandersman, A. (2015). Empowerment evaluation is a systematic way of thinking: A response to Michael Patton empowerment evaluation: knowledge and tools for self-assessment, evaluation capacity building, and accountability. *Evaluation and Program Planning* 52 (2015), 10–14.

Fetterman, D.M., Rodriguez-Campos, L., and Zukoski, A. (2018). *Collaborative, participatory, and empowerment evaluation: stakeholder involvement approaches*. New York: Guilford Publications.

Kirkhart, K.E. (2010). Eyes on the prize: multicultural validity and evaluation theory. *American Journal of Evaluation*, 31(3), 400-413. doi: 10.1177/1098214010373645

Merriam, S.B. (1998). *Qualitative research and case study applications in education*. Jossey-Bass Publishers, San Francisco.

Miller, R.L. (2006). Taking stock of empowerment evaluation: an empirical review. *American Journal of Evaluation*, 27 (3), 296-319. doi: 10.1177/1098214006291015

Ofek, Y. (2017). Evaluating social exclusion interventions in university-community partnerships. *Evaluation and Program Planning*, 60, 46-55. doi: 10.1016/j.evalprogplan.2016.09.004

Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd ed.). Beverly Hills: Sage Publications.

Patton, M.Q. (2002). Two decades of developments in qualitative inquiry: A personal, experiential perspective. *Qualitative Social Work*, 1 (3), 261-283. doi:10.1177/1473325002001003636

Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation, 22*(3), 329-336.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks: Sage Publications.

Patton, M. Q. (2015). The sociological roots of utilization-focused evaluation. *The American Sociologist*, 46, (4) 457-462. doi: 10.1007/s12108-015-9275-8

Robson, C. (2002). *Real world research (2nd ed.)*. Malden: Blackwell Press.

Scriven, M. (2017). Empowerment evaluation 21 years later: There is much to admire about empowerment evaluation. *Evaluation and Program Planning*, 63, 138. doi: 10.1016/j.evalp

Wandersman, A. (2005). Empowerment evaluation. *American Journal of Evaluation*, 26(3), 421-428. doi: 10.1177/1098214005278774

Yin, R. K. (2006). Case study methods. In J. L. Green, G. Camelli, & P.B. Elmore (Eds.). *Handbook of complementary methods in education research* (pp. 111-139). Mahwah, NJ: American Educational Research Association