Experimental ethnographies of early intervention: *a new 'gold standard'?*

Rodolfo Maggio

Narrare i gruppi
Etnografia dell'interazione quotidiana
Prospettive cliniche e sociali, vol. 13, n° 2, dicembre 2018
ISSN: 2281-8960

Rivista semestrale pubblicata on-line dal 2006 - website: www.narrareigruppi.it

Titolo completo dell'articolo	
Experimental ethnographies of early intervention: *a new 'gold standard'?*	
Autore	Ente di appartenenza
Rodolfo Maggio	*University of Oxford*
Pagine 275-298	Pubblicato on-line il 31 dicembre 2018

Cita così l'articolo

Maggio R. (2018). Experimental ethnographies of early intervention: *a new 'gold standard'?* In *Narrare i Gruppi*, vol. 13, n° 2, dicembre 2018, pp. 275-299 - website: www.narrareigruppi.it

research methodology

Experimental ethnographies[1] of early intervention: *a new 'gold standard'?*

Rodolfo Maggio

Abstract

Early intervention in child development has become co-existent with program evaluation by means of randomized controlled trials (RCTs). Although the polarized debate between detractors and promoters of quantitative methods is fading as the idea of interdisciplinarity gains programmatic traction, RCTs are still considered the "gold standard." The application of mixed methods remains limited in evaluation of program effectiveness. In this article I propose three possible forms of integration between ethnography and RCT in the field of early intervention in child development. I argue that such an integration is beneficial for evaluation research and, thus, for the delivery of better early intervention services. In the conclusion, I briefly discuss how drawing ethnography and RCT closer resulted in showing the positive impact of mixing these methods for this kind of evaluative research.

Keywords: early intervention, child development, ethnography, evaluation research, mixed methods, randomized controlled trial, interdisciplinarity.

Etnografie sperimentali dell'intervento precoce: *un nuovo standard di riferimento?*

Riassunto

I servizi di intervento preventivo nello sviluppo infantile sono ormai una cosa sola con la valutazione dei programmi stessi mediante l'utilizzo di studi randomizzati controllati (SRC). Sebbene la polarizzazione del dibattito tra detrattori e promotori dei metodi quantitativi stia svanendo e l'idea d'interdisciplinarità guadagni terreno, gli SRC sono ancora considerati il "gold standard" della valutazione dei programmi. L'applicazione di metodi misti, invece, rimane limitata. In questo articolo propongo tre possibili

[1] The title has been inspired by Sherman and Strang's article "Experimental ethnography: The Marriage of Qualitative and Quantitative Research"(2004).

forme di integrazione tra etnografia e SRC nel campo dell'intervento preventivo nello sviluppo infantile. La mia proposta è che tale integrazione migliora la qualità della valutazione e che, quindi, promuove una migliore fornitura di servizi di intervento preventivo. Nella conclusione, discuto brevemente in che modo avvicinare la mia etnografia a uno SRC abbia dimostrato l'impatto positivo della combinazione di questi metodi per questo tipo di ricerca valutativa.

Parole chiave: intervento preventivo, sviluppo del bambino, etnografia, valutazione, metodi misti, studio randomizzato controllato, interdisciplinarità.

1. *Introduction: Situating Ethnography of Early Intervention*

Early intervention perhaps began with the first home visiting (HV) programs, which Sweet and Appelbaum (2004) date back to the end of the nineteenth century. Home visiting programs seek to respond to the needs of families seen as "at-risk" by means of tailored services and support (Gomby, 2015). Samuel Odom and Mark Wolery locate the development of early childhood intervention at the intersection between special education (Safford, Sargent, and Cook, 1994) and early childhood education (Wolery and Bredekamp, 1994).
Different early-years intervals and definitions of milestone emerged from competing psychological and educational theories, resulting in the development of a diversity of approaches to early intervention. Signaling the need for a synthesis, in 2003 Odom and Wolery advanced a *Unified Theory of Practice in Early Intervention/Early Childhood Special Education*. They wrote that early intervention and early childhood special education is "different from early childhood education in its focus on family-centered services [...], individually planned educational programs, and specialized teaching approaches. It differs from school-age special education in its focus on early developmental skills that are precursors for current and later school success and [...] its emphasis on family" (Odom and Wolery, 2003:164). In addition, early intervention services strive to enhance and support cognitive development, stimulation and language development, including face-to-face interactions with new-borns and infants, social skills, emotional development, as well as diet, physical wellbeing and motor abilities, although different early intervention initiatives might focus particularly on one or more of these aspects (Dalli and White, 2016). Taken all together, the "rationales for studying –and intervening– in early childhood have shifted considerably over time." (Penn, 2016: 475).
Ethnography of early intervention in child development is a new sub-field of ethnographic research that develops at the intersection between ethnography of early childhood and ethnography of education. Summing up the process

that led researchers in education and childhood to look positively at ethnography, Allison James wrote, "Ethnography is becoming the new orthodoxy in childhood research" (James, 2007:246). This quote has been recently re-proposed in *The SAGE Handbook of Early Childhood Research* (Farrell, Kagan, and Tisdall, 2016: 223) in a chapter about ethnography of early childhood. That signals how in-depth, long-term participant observation has been increasingly recognized as a relevant method to study childhood.

Following the differentiation of early intervention from early childhood education, ethnography of early intervention too is to be distinguished from ethnography of early childhood (Konstantoni and Kustatscher, 2016). Typically, early intervention initiatives connect families, service providers, and larger institutional bodies in the common goal of preventing environmental and, thus, developmental risks for children. The 0-3 years interval and the fact that early intervention initiatives are often delivered in areas marked by socio-economic disadvantage considerably delimits the scope of the intervention. That also restricts the focus of the ethnographic observation, although anthropological research tends to connect rather than isolate social phenomena.

While anthropologists of education concentrate on the ways in which children are shaped into context-specific learning and teaching patterns, anthropologists of early childhood are more concerned with what happens before formal schooling begins. Although the study of preschool education can be traced back to the contextualist (LeVine and LeVine, 1966) and ecological (Bronfenbrenner, 1979) traditions, the focus on children between 0 and 3 years of age is a relatively recent development. The anthropological approach gained importance with the study of different models of early education in different cultures and the relationship between the delivery of these models and particular communities (Dahlberg, Moss, and Pence, 1999; Delgado, 2009; New, 1998; Tobin, Vu, and Davidson, 1989; Tobin, Hsueh, and Karasawa, 2009). It is in the context of this diversity of approaches and thematic fields that the ethnography of early intervention is developing.

However, while ethnography has become increasingly important in childhood research, that has not been reflected in increased status in the early intervention industry. The purpose of this article is to reflect on the reasons why this might be and propose a few ways to change that.

First, I synthesize the rationale for using randomized controlled trials (RCTs) as the 'gold standard' in early intervention evaluative research. Second, I present a pragmatic analysis of the methodological compatibilities between ethnography and RCTs in the context of early intervention. Third, I suggest why

an interdisciplinary partnership between RCT and ethnography is not considered the 'gold standard' of evaluation research in early intervention. My overall aim is to propose that an interdisciplinary approach should replace the current orthodoxy in preparation for a comparative theory of early intervention effectiveness that is currently lacking.
My argument is supported by both theoretical reflection and ethnographic research. I conducted 12 months of ethnographic research within Preparing for Life (PFL) an early intervention initiative that pairs new mothers and pregnant women resident in north side Dublin, Ireland, with mentors who influence their behavior in a way that recalls nudge theory (Thaler and Sunstein, 2008) although with a strong emphasis on empathy and social learning theory (Bandura, 1977). My ethnography considers the perspectives and preoccupations of multiple stakeholders, including the mothers, the mentors, as well as the policy makers who design and fund the intervention and the researchers who evaluate it. It should be noted that my collaboration with the researchers who conducted the RCT of this initiative is not presented as exemplary of the proposed argument. My ethnographic fieldwork took place after the completion of the RCT and can be at best considered as a subsequent integration of its data, although it has also inspired the three potential corroborative benefits that, this article proposes, will derive from future integrations.

2. *Evaluating the Effectiveness of Early Intervention*

Primary prevention of life-threatening diseases has long been considered a fundamental right of all children. Recently, a similar kind of logic has been applied to the right to literacy, mental health, and cognitive development. Just like smallpox and measles were considered as diseases that could and should be prevented, so are now psychological pain, social problems, and cognitive delays in infancy as well as lack of preparation to enter the school system (Concha-Eastman, 2016). Hence, both the primary prevention and the early intervention endeavor categorize their fields of application in terms of pathology.
This "therapeutic" attitude (Macintyre, 2013:30-31) at the core of both primary prevention and early intervention is reflected in the epidemiological origin of the standard methodology for the evaluation of their effectiveness, the RCT. In pragmatic terms, RCTs have become the "gold standard" in evaluative research of early intervention (Stewart-Brown et al., 2011) because governments and agencies prefer to finance only those intervention programs that are evidenced to produce the best results. If it is possible to isolate the specific factors that

made a particular intervention effective, the epidemiological discourse implies, it is also possible to predict the outcomes of an investment of taxpayers' money. It then becomes possible to deem such investment as profitable and beneficial or not.

Such evaluative conclusions rest on methodological grounds. "The main appeal of the RCT comes from its potential to reduce selection bias. Randomization, if done properly, can keep study groups as similar as possible at the outset, so that the investigators can isolate and quantify the effect of the interventions they are studying. No other study design gives us the power to balance unknown prognostic factors at baseline." (Jadad and Enkin, 2008: 29). Research participants are assigned to either a treatment or a control group and, in the case of early intervention in child development, the former group will receive the intervention (in the form of, for example, a number of home visits, parenting classes, and educational materials). The latter group will receive a lower 'dose' of the treatment, or no treatment at all. At the end of the experiment, the differences between the two groups will be neutralized in the randomization process and all remaining differences will be attributed to the effect of the intervention.

Notwithstanding the methodological appeal of the RCT, its ability to evaluate the effectiveness of early intervention programs has been repeatedly called into question. Among the most common criticisms, the following three are noteworthy.

First, although participants are randomly assigned to different study groups and, as a consequence, the study is more likely to be free from allocation bias, random allocation poses no solution to other important biases. Among the biases that could be introduced along with an RCT methodology, it is possible to list selection bias, ascertainment bias before and after data collection, and other kinds of bias (Jadad et al., 2008). Among the biases relevant for the present discussion, intervention choice bias is of particular interest. Intervention choice bias depends on the kind of intervention that has been selected for a particular population. It occurs when the way the intervention works *per se* influences the data and the data collection process.

For example, if the intervention is not expected to produce major results in the early stages of the program, it is necessary to wait for the effects to become larger and thus easier to collect. In the case of an early intervention program, that time might be relatively long. Participants are expected to take up new parenting practices throughout the program, and to have changed their patterns of behavior towards its end. Before such changes become observable,

even more time might be necessary. It is therefore of critical importance that the data collection is carried out at the time when the particular kind of intervention is expected to produce the relevant results. Selecting a particular point in time, therefore, depends upon very specific contextual variables.

The second criticism considers self-report as one of the weaknesses of the RCT methodology. In the case of early intervention with parenting behavior, self-report questionnaires and interviews about a parent's attitudes to parenting are often biased (Milner and Crouch, 1997; Straus et al., 1998). For example, sometimes mothers present themselves as better parents compared to what their children and partners say of them (Bögels and Melick, 2004). Nevertheless, self-report is almost ubiquitous in the evaluation of early intervention initiatives (Janus and Offord, 2007), particularly because of the supposed absence of researcher interference.

A third, more general, problem arises from the experimental character of the RCT. The fact that a functional relationship can be established between a treatment and an effect by means of a statistical correlation is not sufficient to conclude that the said treatment causes the said effect. While there might be some kind of relationship between baseline and outcome data for specific measures, there might as well be other, perhaps important but unobserved factors connecting the treatment and the effects. If these factors are not examined in detail and the statistical relation is accepted as the sole or primary expression of the causative relationship between the treatment and the effect, such a finding would have a limited range of application. In particular, it would not be possible to use it to support the argument that a particular component of the intervention has been effective. The effective element might be locatable in an unobserved factor, one that was perhaps co-existent or co-located with the observed variable, but not captured by the research method.

That is not to say that the relationship captured by the RCT is not relevant. It is to say that it is necessary to exclude other factors as not accountable for the observed effects and to indicate a precise reason for doing so. For, the "core of the scientific method is not experimentation per se, but the strategy connoted by the phrase plausible rival hypotheses" (Campbell, 1994: ix). While it is arguable that a given treatment caused a particular effect, in the absence of a close examination of rival hypotheses it is difficult to verify whether that argument is the most valid. But the list of plausible rival hypothesis cannot be completed unless a thorough examination of the context where the program is delivered over a relatively long period of time. Without a list of plausible rival hypotheses and alternative explanations, that verification is not possible.

These are just three out of many and more subtle criticisms that have been formulated against the RCT methodology. There is an extensive literature on the subject. To quote one of the most popular textbooks in research methods in education, RCTs are deemed to belong to "a discredited view of science as positivism" (Cohen, Manion, and Morrison, 2013: 318). Although criticisms of this kind are common in educational research today, it is equally common for RCTs to be used alongside early intervention initiatives. Such a contradiction has been explained as a discrepancy of standards between different stakeholders (Glasgow, Lichtenstein, and Marcus, 2003) such as scholars and policy makers. The following section suggests a few ways in which this discrepancy can be de-emphasized in favor of a more collaborative attitude between researchers from different methodological backgrounds and other stakeholders in the early intervention sector.

3. *Interdisciplinary, Problem-Specific, Collaborative Methods*

Notwithstanding the controversies outlined above, most early intervention practitioners and their evaluators adopt the view that an RCT can tell us whether a treatment has been effective or not on the basis of what is considered logical reasoning and correctness of argumentation. However, even accepting the logic that underpins the RCT, on its basis it is not possible to understand the process of differentiation of treatment and control group. The RCT is not designed to illustrate that process. It only relies on observations conducted before and after the intervention is delivered, rather than on continuous observations throughout the program. Although in some cases RCTs are conducted at several points throughout the intervention, such as at 6 months intervals, that is mostly not the case. And even if that was the case, at-interval surveys are no substitute for long-term observation.

Ethnography is constituted precisely by long-term, in-depth, participant observation. The logic that convinces ethnographers of the trustworthiness of their method is that any phenomenon taking place within a context is embedded within that context and must be understood from the point of view of those who are most familiar with said context (Jessor and Shweder, 1996). In order to do that, it is necessary to simultaneously observe and participate in the everyday life of a group of people for a relatively long time (Malinowski, 1922). Within the context of an early intervention initiative, participant observation allows the researcher to describe the process through which changes brought in by an exogenous phenomenon, such as a parenting program, are converted

into actual practices, if at all. While the logic of ethnography cannot tell whether an intervention has been effective or not in statistical terms, it can illustrate the process that resulted in the outcome of the intervention. In other words, while it cannot measure the extent of the change occurred, it seeks to explain why and how the intervention worked the way it did.

In sum, the RCT can measure whether the control group has changed or not, but cannot explain why, whereas ethnography can illustrate how that result was achieved, but cannot estimate its extent. In answering these two different yet interrelated sets of questions, the rationales underpinning RCT and ethnography appear as both different and logically compatible. As such, they can be used jointly to understand interventions much better than if they were applied in isolation.

This is a purely theoretical argument. Hence it is necessary to explain how the combination of these methodologies can provide some form of practical advantage for the stakeholders of an early intervention initiative. I mention three ways in which the logical compatibility between ethnography and RCT can be concretized.

First, ethnographers can produce stark descriptions at the early stages of the program when the kind of data necessary for the RCT cannot be collected. As mentioned in the previous section, it is necessary to tailor the RCT to the intervention timeline in order to avoid intervention choice bias. At the early stages of early intervention programs, changes in the practices of recruited participants can be subtle and slow. This is particularly the case when a relationship of trust must develops between the mentor and the families for the program to be delivered effectively. Changes in parenting skills and their consequences on child development might take months, even years to yield statistical results. Thus, capturing changes with quantitative instruments can be hard when these effects are statistically small. Tasking ethnographers with the observation of changes at the early stages of the intervention can be helpful here. Ethnography can provide stark illustrations of individual instances of change at the early stages of the study-intervention, or the lack thereof. Furthermore, these can be used to identify unwanted circumstances early and take corrective action.

Second, in-depth, long-term participant observation can be used to improve the accuracy of survey methods (Sieber, 1973). As mentioned earlier, self-report questionnaires can be a source of bias. In addition to the bias introduced by the research participants in answering questions, the survey in itself can be methodologically problematic. When experimental and quasi-experimental approaches lack a prospective qualitative component, surveys on-

ly ask questions from the perspective of outsiders. Questions formulated in this way risk resonating very little with the research participants (Hymes, 2003). For this reason, surveys should be informed by qualitative data collected before the RCT starts. On the basis of qualitative interviews and focus groups, it is possible to ask questions that, while being relevant for the research agenda of the outsiders, are translated into terms that resonate with the insiders. The more the questions resonate well with the research participants, the more they will feel represented by their own answers.

While the translation of questionnaires in terms understandable by the target population is a relatively established routine, understanding the current concerns of the research participants is less common. If the survey is concerned with matters that do not correspond to the participants' perspectives, the RCT will not evaluate important aspects of the intervention. For example, according to the RCT conducted alongside PFL, the recruited families rated the importance of the mentor-mother relationship "very highly." My ethnography confirms that the quality of the relationship is a crucial ingredient for the effective delivery of the program. That means that the relationship has to be "good" in order for the program to be effectively delivered. The RCT, though, was not designed to understand this "good relationship", neither from the point of view of ensuring it nor from the point of view of measuring its impact. That substantially limits the ability of the researchers to understand how the intervention worked and which components are responsible for its effectiveness, and why.

If in-depth participant observation is conducted within the concerned group before the survey is designed, it will be possible to complement the questionnaire with questions that could not be deemed relevant with less intensive qualitative research methods. While it is important to translate the questions relevant for the outsiders in terms that are meaningful for the insiders, it is equally important to value what is relevant for the insiders and translate it into questions that the outsiders can learn to treat as meaningful.

Third, the blend of ethnography and RCT balances the experimental ethos of the latter with the participatory and reflexive ethos of the former. The genealogy of the RCT can be traced back to the experimental paradigm of epidemiological research, as mentioned above. For this reason, its logic depends on theoretical rigor often so rigid that it can hardly be applied in contexts marked by socio-cultural complexities. All human settings are complex and cannot be controlled as if their parts could be isolated from each other. For example, within a community of interconnected service recipients it is impossible to en-

tirely avoid contamination, a term with clear epidemiological derivation that is used to express the unintended transmission of some of the intervention from the treatment to the control group.

The main challenge for the application of a methodology rooted in the experimental paradigm in a non-experimental setting is that single causes cannot be methodologically isolated because they are not isolated from each other substantially. In that sense, an RCT not compromised by the features of the context is not possible. The bar, still, is set so high that even the smallest difference from optimal experimental conditions might be considered to invalidate the RCT as a whole. But since experimental conditions cannot be entirely satisfied in complex human settings, it is necessary to accept less than optimal conditions. That, however, can only be done as long as a clear methodology is in place to compensate for the bias that less-than-optimal conditions introduce.

For example, it suffices that one person drops out of the study for the randomization to be spoiled and the theoretical validity of the experiment to be hampered. There has never been an RCT without human dropouts, so this problem concerns all trials, not only those designed alongside early intervention initiatives. Dropouts undermine the reliability of the experiment because it is not possible to compare baseline data with the data they would have put in had they remained in the study. Compensating for the absence of dropouts, inevitably, requires the introduction of statistical strategies the choice of which is largely subjective.

In general, the missing data of the dropouts can be compensated with a variety of statistical remedies. These range from limiting the analysis to non-missing data, to inverse probability weighting (IPW). However, these strategies can be questioned on methodological grounds. For example, IPW can compensate missing data only as long as enough information is available about the entire population to predict the probability of non-missingness. However, the concept of "enough information" leaves room for interpretation and, again, the introduction of subjective bias.

If an evaluative study incorporates values that are not quantifiable, develops reflexivity, and relies less on the experimental character of the investigation, these problems might seem less of a concern. For example, rather than using statistics to compensate for the missing data of the dropouts, RCT researchers might task ethnographers to concentrate, in a participative way, on individual research participants that, according to the predictive models, are more likely to leave the study. Ethnographers can discuss with these participants without necessarily touching on the issue of abandoning the study, but the fact of

providing them with one-to-one care and attention might in itself decrease the likelihood of their dropping out. Control group research participants are often more likely to disengage because they feel less important than the treatment group, even in the case of a double-blind RCT. As they are members of the same community of the treatment group members, some of them would eventually realize that they are not receiving as much support or 'dose' as other research participants. Still, they are a crucial component of the RCT, even if they might not perceive that to be the case. If they drop out it would be just as difficult for the researchers to accurately evaluate the effectiveness of the intervention as if members of the treatment group had abandoned. The problem is that their importance is not reflected in the engagement of the interventionists with them. If, however, they were treated as a source of research insights as much as the members of the treatment group, they would understand better how important they are. In order to make that happen, the ethnographer can engage with them through participant observation, thereby limiting the number of dropouts.

These are just a few examples of how the methodological compatibility between ethnography and RCT could be converted into the pragmatics of interdisciplinary evaluation research. It is clear that both ethnographers and RCT experts would benefit greatly from this interdisciplinary, problem-specific collaboration. It follows that it is for reasons other than methodological and theoretical convenience that this kind of collaboration has not yet gained in popularity. In the following sections I briefly discuss what these other reasons might be.

4. *Reasons for a lack*

Given the above argument, it is not clear why the "gold standard' is not to design evaluation research in early intervention as a partnership between ethnographers and RCT experts. Four main reasons might explain the limited application of interdisciplinarity in evaluative research on early intervention.

First, the ethics of applying an intervention model to change the behavior of people regarded as in need of social support is problematic. Many ethnographers, as well as experts from related disciplines, consider ethnography antithetical to the intervention endeavor. Often they are openly against the application of an external perspective onto a group of people, however mediated by ethical considerations. They think that getting closer to the point of view of the

research participants/service recipients in order to ensure compliance with an externally developed program should not be the purpose of their discipline.

Second, early intervention evaluative research inherits the historical opposition between quantitative and qualitative research methods. The Global Research Council Report (GRC) 2016 states that, according to the "global literature" interdisciplinarity "has a key role to play in addressing the grand challenges that society faces" (Gleed and Marchant, 2016: 5). Yet, the Report recognizes that interdisciplinarity remains essentially a good idea rather than a common practice. The reasons are many and complex but they can perhaps be synthetized in the fact that interdisciplinarity is a relatively recent development. Although it is possible to identify instances of interdisciplinarity before the 1970s, these were generally related to the exceptional intellect of isolated figures. Interdisciplinarity as we know it today only started to gain traction in the 1970s; in the context of social sciences this happened even later (Small, 2011; Tashakkori and Teddlie, 1998; Tashakkori and Creswell, 2007; Johnson and Onwuegbuzie, 2004; Ragin, 2014; Bollen and Paxton, 1998).

Third, there is still limited familiarity between qualitative and quantitative researchers. Stephen Bell and Peter Aggleton reflected on the consequences of this lack of familiarity. (Bell and Aggleton, 2012). In their edited volume, *Monitoring and Evaluation in Health and Social Development* (2016) they argue in favor of ethnographic methods in evaluative research. However, they do not provide an example of how an integration with RCTs can be designed. In the second part of the volume the contributors do address the theme of research design, but a fundamental opposition persists between qualitative/quantitative, emic/etic, deductive/inductive, experimental/exploratory, and so on. In that way, each approach remains fundamentally separate from the other rather than integrated with the other.

Quantitative researchers are all too often caricaturized as cold, ultra-rationalistic machine-like beings who believe that an objective reality can be perfectly known by means of increasingly sophisticate heuristic models (Saracho, 2016:15). Qualitative methodologists consider this approach inappropriate because it is designed in isolation from the context where it is to be applied. As a consequence, the quantitative representation of early intervention programs is sometimes considered inaccurate, but also ethically questionable, as mentioned earlier, because the superimposition of an external perspective

homogenizes difference and obliterates self-representation[2]. Another negative consequence of experimental detachment is that the research insights it produces are so distant from the every-day lives of practitioners and service recipients that they have little application and social relevance (Foster and Mash, 1999).

On the other hand, qualitative approaches like ethnography are often assumed not to be representative of the wider social reality in which a small group is embedded, essentially because they do not rely on methods measuring the extent to which the given group differs in significant ways from the broader population. Another reason why there might be diffidence is that the kind of data that ethnographers collect often require a narrative form to be presented, one that does not lend itself to secondary analysis. All in all, while interdisciplinarity gains popularity, at least in some disciplines, the reciprocal diffidence between qualitative and quantitative researchers limits the development of interdisciplinary practices.

While academic debates are often instigated as means of building some careers and destroying others, the opposition between quantitative and qualitative approaches is perhaps grounded in the genuine attempt to generate more suitable means to understand social phenomena. However, their heuristic reach remains limited because complex evaluation questions are still addressed as if the phenomena they intend to understand could be broken down into qualitative and quantitative components that could be studied in isolation from each other, as opposed to an interdisciplinary approach constituted by a complex and concrete integration of qualitative and quantitative methods.

Finally, the fourth reason relates to the availability and distribution of funding. Agencies are committed to fund, when possible, only programs that are deemed or expected to be effective according to an evidence-based paradigm. That has the unintended consequence of encouraging a methodological approach "that flattens complex change processes into overly simple causes and effects" (Bell and Aggleton, 2016: 4) rather than approaches that describe the complexities linking specific causes to their consequences. More generally, according to the GRC, interdisciplinary research receives less funding because peer review procedures are not designed to evaluate interdisciplinary proposals. "This is partly due to a lack of reviewers who understand how to evaluate interdisciplinary research, and the related circular problem that there is a need to

[2] A reflection on how to manage these biases would benefit from the literature on reflexivity in ethnography (Davies, 2012), especially when applied to service delivery (see, for example, Pellatt, 2003).

expose more reviewers to interdisciplinary projects"(Gleed and Marchant, 2016). It might be the case that similar circular reasons explain why early intervention initiatives are not designed alongside research partnerships that include RCTs and ethnography as compatible research methods.

5. *Drawing ethnography and RCT closer*

In this article I argued that one of the main reasons why the benefits of a collaboration between RCT experts and ethnographers are not realized is that these benefits are not readily apparent for both scientists and policy makers. However, there are exceptions to this general claim. For example, Carlos Moedas once said:

"I would argue that the job of scientific advisor has dramatically changed. I think that the scientific advisor is no longer the expert who provides the answers. […] It's about the process of collecting evidence in a multi-disciplinary world. […] People will only accept the answers if they understand the process" (Moedas, 2016).

Moedas' words seem to suggest that the argument proposed in this article might not be an isolated attempt to formulate a different way of doing evaluative research. A case in point is provided by Martin Walsh, who examined "Oxfam's use of interpretive research to deepen the findings of project evaluations based on the use of quantitative survey methods" and claimed that "Oxfam's experience suggests that the intersubjective and interpretive modes of enquiry that characterize ethnographic research […] can be integrated with experimental approaches in order to generate more effective learning for programs." (Walsh 2016: 219). It appears therefore that instances of politicians and agencies understanding the value of interdisciplinary evaluation, as opposed to the much celebrated "gold standard", can be identified.
In academia, instances of this new trend can be found, for example in a job description recently advertised on the website of The UCD School of Education. The document briefly describes a "mixed methods evaluation of an intervention […] designed to support children's literacy, their rights and well-being. One researcher will be mainly responsible for managing the day-to-day running of a large-scale, cluster-randomised controlled trial whereas the other researcher will be mainly responsible for a longitudinal in-depth qualitative research study into the everyday lives of a subsample of children and their families."

My own fieldwork in Dublin can be considered another instance in which ethnographic methods and RCT have been drawn closer in order to understand a phenomenon more deeply and holistically. Although the two methodologies have been used independently as part of two institutionally separated projects, collaborations between different stakeholders (both the researchers and research participants) resulted in a series of valuable exchanges that reveal a growing enthusiasm about mixing RCT and ethnography. In this concluding section, I wish to explain in more detail how the two methodologies have been drawn closer. In order to do so, I will briefly describe the evaluation process, its results, and how these were integrated with ethnographic data.

As part of the RCT of PFL, the Evaluation Team led by Orla Doyle collected data by means of the following methods: questionnaires, observations, and direct assessments when the children reached 6, 12, 18, 24, 36, and 48 months of age; the analysis of hospital records pertaining to both mothers and children; an online survey about the children's school readiness, completed by their teachers; and a series of interviews conducted with the PFL children about their experiences of school life. The results suggested that the PFL program improved the cognitive development of children in the treatment group from 18 months onward.

More specifically, at the time when they entered the school system, they had "better general cognitive functioning, spatial abilities, non-verbal reasoning skills, and basic numeracy skills" (PFL Evaluation Team, 2016: xv) than the control group. At this same stage, the PFL program had a profound impact on the overall verbal abilities of the high-treatment group children, with positive consequences on their speaking, listening, reading and writing skills. They were also slightly more keen to explore and learn, and much less prone to hyperactivity and inattentive behavior in the classroom. However, the "program had no impact on children's aggression, oppositional-defiance, anxious behavior, or on their prosocial, respectful behaviors according to the teacher reports" (xvii). Still, children in the high-treatment group used the hospital services significantly less and less likely for emergency reasons. Their fine and gross motor skills also scored better than the control group. Overall, the Evaluation Team found that the children in the high-treatment group were significantly more prepared to enter the school system than the control group, and quantified the cognitive gap between them in 10 IQ points.

Although these results seem to indicate that the PFL program was generally successful, this is only partially true. The program worked in the sense that it improved the preparation of children to enter the school system. However, the

logic model was inaccurate in predicting *how* these results were going to be achieved. Even if the program did cause the hypothesized effects in children, these were not linked to effects on their parents, that is, according to the initial hypothesis.

The prediction was, indeed, that a positive change in parents' wellbeing, behaviors, and attitudes about children and parenting would take place because of the program, and that this would stimulate the changes in children. The data collection and analysis, though, demonstrated that children changes went in the hypothesized direction even if the same was not observed in the parents. According to the RCT, the parents did not really change their attitudes and wellbeing throughout the program, nor at school entry stage, or at least not in the measures that were collected (self-esteem, self-efficacy, parenting beliefs, parenting attitudes, parenting styles). They did change their behaviors, but that was not influential enough to cause a change in their children. It follows that, while the methodology -that is, the connection between the data collection and the research question- was sound, the theoretical framework used to conceptualize the internal mechanisms of change was, at least partially, inaccurate.

That does not mean that the Evaluation Team did not observe any change in the parents. According to the Final Report, parents changed their behaviors in ways that might be linked to their children's positive changes. Parents spent more time interacting with their children during infancy, less time watching TV in toddlerhood, and had more organized routines, which might have had a positive impact on cognitive and language development. The high-treatment children had better physical health and wellbeing, which might be attributed to the minor exposure to cigarette smoke, a healthier diet, and timely immunization.

The study also found changes in parents' wellbeing. Higher levels of happiness for the day overall were identified using a Day Reconstruction Method, both when the parents were with and without the PFL child. This last measurement might be explained by the fact that the PFL program encouraged the parents to become aware of what they do, thereby calming anxieties related to not doing enough. Also, Doyle et al. proposed that the increased parenting effort resulting from the fact of being enrolled made parents more appreciative of the time not spent with their children. As for the time spent with the child, Doyle et al. argue that although there were "no differences in the amount of time participants spend with their children in either group, the results suggest that the higher positive affect experienced by the treatment group may be driven by differences in the quality of the episodes rather than the quantity of episodes." The driver of positive change was, in other words, the quality rather that the

quantity of time. However, no treatment effects were found for the negative aspects of well-being.

The overall limited impact of the PFL program on the wellbeing of the participant parents might be partially explained by the prevalence of diagnosed mental health conditions in the community, which has been observed at baseline. While parents with diagnosed mental health issues were in need of therapeutic care, the PFL mentors were not acting in a way that might be believed to provide that. It follows that the prevalence of unresolved mental health issues should have been taken into consideration by the logic model, especially considering the existing evidence that HV programs tend to be less effective with participants experiencing mental health issues (Ammerman, Putnam, Bosse, Teeters, & Van Ginkel, 2010; Fergusson, Horwood, & Grant, 1998; Sweet & Appelbaum, 2004).

The PFL evaluation team has been the first to examine "the utility effects of a targeted early intervention program using multiple measures of well-being" (Doyle et al., 2017: 17), although the relevance of the aggregated measure of experienced affect to evaluate the effectiveness of policies had already been proposed. Instead, the pre-existing literature highlighted the lack of treatment effects on *negative* measures of well-being. Doyle et al. found that PFL's lack of effect in ameliorating negative emotional states was in line with the previous evaluative studies on the effectiveness of HV programs. However, they also demonstrated some dimensions of positive effect, which challenged the widely held assumption in previous systematic reviews on the effectiveness of HV programs. Nevertheless, the process through which the effectiveness was achieved remained unclear.

In sum, while some changes were observed in parent's behaviors and a limited positive impact was observed on their wellbeing, no substantial changes were observed in their parental attitudes. Cumulatively, these findings invalidate the "logic model." The logic model, within the theoretical framework of the PFL program, incorporates three psychological theories of development: the theory of human attachment (Bowlby, 1969), the socio-ecological theory of development (Bronfenbrenner, 1979), and the theory of social learning (Bandura, 1977). These theoretical axes form the basis of a logic model conceived as the basic mechanism to tackle the intergenerational transmission of disadvantage in children's health and cognitive, behavioral, and emotional development (Najman *et al.,* 2004; Shonkoff & Phillips, 2000). The model hypothesizes that school readiness can be increased by means of an intervention on parents, rather than children directly. It focuses particularly on knowledge, attitudes and

feelings, but it is also intended to improve parenting behavior and, by consequence, child behavior. The hypothesis has been formulated on the basis of other studies suggesting the inherent logic of the model reflected the internal mechanism of the intervention.

The question[3] then is, why did PFL improve children's school readiness? The final report of the Evaluation Team recognizes that "standardized measures and statistical tests can miss the small, but important, changes that parents have made as a result of the intervention."(82) Hence, although parents' well-being and parenting attitudes did not seem to be influenced by the program, this impression might be caused by the methods applied to capture the hypothesized change. Here comes the interest for a methodology that mixes the RCT with other methods.

The methodology used to establish the effectiveness of the PFL Program is better suited to do exactly that than to identify the reason why such effect was produced. The answer to the "why question" requires a close examination of the external and internal mechanisms of the intervention, a research activity that is substantially different from the data collection routine and data analysis of the RCT. As mentioned above, RCTs are sometimes conducted at different points in time, but a long-term observation would be necessary in this case. In order to observe how the PFL staff worked before, during, and after the intervention (the external mechanisms) and how the participants responded (the internal mechanisms) it would have been necessary to have an embedded observer for an extended period of time.

In this article, I argued that, in order to answer the "why question", the methodological specificity of the RCT can be complemented with an ethnographic study. Within the context of an ECI initiative, participant observation allows the researcher to describe the process through which changes brought in by an exogenous phenomenon, such as a parenting program, are converted into actual practices, if at all. Although it is more difficult to demonstrate the effectiveness of a parenting program with ethnographic methods than with an RCT, an ethnographer might be better positioned to observe and describe the process leading to the outcome of the intervention. In other words, while the ethnographer cannot measure the occurred changes, he or she can establish an explanatory description of why and how the intervention resulted in the outcome values established with an RCT. The RCT, in turn, can only provide an explanation of those results on the basis of the logic model, which in some cases can be contradicted by the results themselves. It follows that ethnogra-

[3] This question is addressed in depth in a dedicated publication, currently under review.

phy and RCT are two research methods that can answer two different but interrelated research questions in a methodologically sound way. Just like these two questions, the logics underpinning RCT and ethnography also appear as different but compatible. It follows, in conclusion, that they should be used in conjunction, rather than in isolation, to evaluate this kind of interventions.

References

Ammerman, Robert T., Frank W. Putnam, Nicole R. Bosse, Angelique R. Teeters, and Judith B. Van Ginkel. (2010). "Maternal Depression in Home Visitation: A Systematic Review." Aggression and Violent Behavior 15(3): 191–200.

Bandura, A. (1977). *Social Learning Theory*. Oxford, England: Prentice-Hall.

Bell, S., Aggleton, P., (2012). Integrating ethnographic principles in NGO monitoring and impact evaluation. *Journal of International Development* 24(6):795-807.

Bell, S., Aggerton, P. (2016). *Monitoring and Evaluation in Health and Social Development: Interpretive and Ethnographic Perspectives*. London: Routledge.

Bögels, S.M., van Melick, M. (2004). The relationship between child report, parent self-report, and partner report of perceived parental rearing behaviors and anxiety in children and parents. *Personality and Individual Differences* 37(8): 1583–1596.

Bollen, K., Pamela, P. (1998). Detection and Determinants of Bias in Subjective Measures. *American Sociological Review* 63(3): 465–78.

Bowlby, J. (1969). *Attachment and Loss. 1. Attachment*. New York: Basic Books.

Bronfenbrenner, U. (1979). *The Ecology of Human Development*. Cambridge, Massachusetts, and London, England: Harvard University Press.

Campbell, D. (1994). Retrospective and prospective on program impact assessment. *Evaluation Practice* 15(3): 291-298.

Concha-Eastman, A. (2016). "Introduction: Home visitation as a primary prevention tool for violence." In: Roggman, L., Cardia, N (eds.) *Home Visitation Programs*. Springer International Publishing, pp. 3-8.

Dahlberg, G., Moss, P., Pence, A.R. (1999). *Beyond Quality in Early Childhood Education and Care: Postmodern Perspectives*. London: Falmer/Routledge.

Dalli, C., White, J. (2016). Group-based early childhood education and care for under-2-year-olds: Quality debates, pedagogy and lived experience. In Farrell A,

Kagan SL, and Tisdall EKM (eds.), *The SAGE handbook of early childhood research*. London: SAGE, pp. 36-50.

Davies, C. A. (2012). *Reflexive ethnography: A guide to researching selves and others*. Routledge.

Delgado, M.A. (2009). *Community Involvement in Services for Young Children: Accepting, Resisting and Proposing Alternatives to Mainstream Views*. Saarbrucken: VDM.

Doyle, Orla, Liam Delaney, Christine O'Farrelly, Nick Fitzpatrick, and Michael Daly. (2017). "Can Early Intervention Improve Maternal Well-Being? Evidence from a Randomized Controlled Trial." PLOS ONE 12(1).

Farrell, A., Kagan, S.L., Tisdall, E.K.M. (2016). *The SAGE Handbook of Early Childhood Research*. London: SAGE.

Fergusson, David M., Hildegard Grant, L. John Horwood, and Elizabeth M. Ridder. (2005). "Randomized Trial of the Early Start Program of Home Visitation." Pediatrics 116(6): 803–809.

Foster, S.L., Mash, E.J. (1999). Assessing social validity in clinical treatment research: Issues and procedures. *Journal of consulting and clinical psychology* 67(3):308.

Gomby, D. (2005). *Home visitation in 2005: Outcomes for children and parents. Invest in kids.* Committee for economic development invest in kids working group. http://legis.wisconsin.gov/lc/committees/study/2008/SFAM08/files/GombyHVoutcomes2005.pdf

Glasgow, R.E., Lichtenstein, E., Alfred, C.M. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health* 93(8):1261-1267.

Gleed A. and Marchant D. (2016), *Interdisciplinarity: Survey Report for the Global Research Council*. Research Councils UK and the Science and Engineering Research Board. Retrieved August 15, 2017. http://www.globalresearchcouncil.org/sites/default/files/pdfs/Interdisciplinarity%20Report%20for%20GRC_DJS%20Research.pdf.

Hymes, D. (2003). *Ethnography, linguistics, narrative inequality: Toward an understanding of voice*. Taylor & Francis.

Jadad, A.R. and Enkin M.W. (2008), Bias in randomized controlled trials. In: Jadad AR and Enkin MW (eds.) *Randomized Controlled Trials: Questions, Answers, and Musings, Second Edition*, edited by. Oxford: Blackwell, pp. 29-47.

James, A. (2007), Ethnography in the study of children and childhood. In: Atkinson P, Coffey A, Delamont S, Lofland J and Lofland L (eds) *Handbook of Ethnography*. London: SAGE, pp. 246-58.

Janus M. and Offord D.R. (2007), Development and psychometric properties of the early development instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioural Science* 39:1–22.

Jessor, R., Colby, A., & Shweder, R. A. (Eds.). (1996). *Ethnography and human development: Context and meaning in social inquiry*. University of Chicago Press.

Johnson, R.B. and Onwuegbuzie A.J. (2004), Mixed methods research: A research paradigm whose time has come. *Educational researcher* 33(7):14-26.

Konstantoni, K. and Kustatscher, M. (2016), Conducting Ethnographic Research in Early Childhood Research: Questions of Participation. In: Farrell A, Kagan SL, and Tisdall EKM (eds.) *The SAGE handbook of early childhood research*. London: SAGE, pp. 223-240.

LeVine R.A, Bloom Lloyd LeVine, B. (1966), *Nyansongo: a Gusii community in Kenya*. New York: Krieger.

MacIntyre, A. (2013), *After virtue*. London: Bloomsbury.

Malinowski, Bronislaw. (1922). *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea*. London, UK: Routledge.

Milner, J.S. and Crouch J.L. (1997), Impact and detection of response distortions on parenting measures used to assess risk for child physical abuse. *Journal of Personality Assessment* 69:633–650.

Moedas, C. (2016), "No title". Brussels: International Network for Government Science Advice Conference. Retrieved on August 15, 2017. https://ec.europa.eu/commission/commissioners/2014-2019/moedas/announcements/international-network-government-science-advice-conference_en

Najman, Jake M., Rosemary Aird, William Bor, Michael O'Callaghan, Gail M. Williams, and Gregory J. Shuttlewood. 2004. "The Generational Transmission of Socioeconomic Inequalities in Child Cognitive Development and Emotional Health." *Social Science and Medicine* 58(6): 1147–58

New, R.S. (1998), Theory and Praxis in Reggio Emilia: They Know What They are Doing and Why. In: Edwards C, Gandini L, and Forman G (eds.) *The Hundred*

Languages of Children: The Reggio Emilia Approach – Advanced Reflections, NJ: Ablex, pp. 261-284.

Odom, S., Wolery, M (2003) A unified theory of practice in early intervention/early childhood special education evidence-based practices. *The Journal of Special Education* 37(3):164-173.

Pellatt, G. (2003). Ethnography and reflexivity: emotions and feelings in fieldwork. *Nurse Researcher (through 2013)*, *10*(3), 28.

Penn, H. (2016). Social and Political Landscapes of Childhood. In: Farrell A, Kagan SL, and Tisdall EKM (eds.) *The SAGE handbook of early childhood research*. London: SAGE, pp. 469-485.

PFL Evaluation Team. (2016). "Preparing for Life: Early Childhood Intervention. Final Report. Did Preparing for Life Improve Children's School Readiness?" University College Dublin. Geary Institute for Public Policy. Ragin C.C. (2014). *The comparative method: Moving beyond qualitative and quantitative strategies*. Oakland, CA: University of California Press.

Safford, P., Sargent, M., Cook, C. (1994), Instructional models in early childhood special education: Origins, issues, and trends. In Safford P, Sargent M and Cook C (eds.) *Early childhood special education: Yearbook in early childhood education. Volume 5*. New York: Teachers College Press. pp. 96-117.

Saracho, O.N. (2017). Writing and Publishing Qualitative Studies in Early Childhood Education. *Early Childhood Education Journal* 45(1):15-26.

Sherman, L.W., Strang, H. (2004), Experimental Ethnography: The Marriage of Qualitative and Quantitative Research. *The Annals of the American Academy of Political and Social Science* 595: 204–22.

Shonkoff, J., Phillips, D. (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. 2000. Washington, DC: National Academies Press.

Sieber, S. D. (1973). The integration of fieldwork and survey methods. *American journal of sociology*, *78*(6), 1335-1359.

Small, M.L. (2011). How to Conduct a Mixed Methods Study: Recent Trends in a Rapidly Growing Literature. *Annual Review of Sociology* 37(1):57– 86.

Stewart-Brown, S., Anthony, R., Wilson, L., Winstanley, S., Stallard, N., Snooks, H., Simkiss D. (2011). Should randomized controlled trials be the "gold standard" for research on preventive interventions for children? *Journal of Children's Services* 6(4):228-235.

Straus, M.A., Hamby, S.L., Finkelhor, D., Moore, D.W., Runyan, D. (1998). Identification of child maltreatment with the Parent–Child Conflict Tactics Scales: Development and psychometric data for a national sample of American parents. *Child Abuse & Neglect*, 22(4):249–270.

Sweet, M.A., Appelbaum, M.I. (2004). Is home visiting an effective strategy? A meta- analytic review of home visiting programs for families with young children. *Child development*, 75(5), pp.1435-1456.

Tashakkori, A., Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. London: SAGE.

Tashakkori, A., Creswell, J.W. (2007). Editorial: The New Era of Mixed Methods. *Journal of Mixed Methods Research* (1):3-7.

Thaler, R.H., Sunstein, C.R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven & London: Yale University Press.

Tobin, J.J., Wu, D.Y.H., Davidson, D.H. (1989). *Preschool in Three Cultures*. New Haven, CT: Yale University Press.

Tobin, J.J., Hsueh, Y., Karasawa, M. (2009). *Preschool in Three Cultures Revisited: China, Japan and the United States*. London: University of Chicago Press.

Walsh, M. (2016), Using interpretive research to make quantitative evaluation more effective. In: Bell S and Aggerton P (eds.) *Monitoring and Evaluation in Health and Social Development: Interpretive and Ethnographic Perspectives*. London: Routledge. pp. 219-231.

Wolery, M., Bredekamp, S. (1994). Developmentally appropriate practice and young children with disabilities: Contextual issues in the discussion. *Journal of Early Intervention* 18:331–341.