

STRENGTHENING EVIDENCE IN PUBLIC POLICY EVALUATION: ETHICAL AND METHODOLOGICAL CHALLENGES IN SOCIAL IMPACT MEASUREMENT AND THE ROLE OF PSYCHOLOGY

Rui Serôdio¹, Alexandra Serra^{1,2}, Sofia Brito¹, Sofia Matias¹, Catarina Mano¹,
& Sandra Nogueira¹

¹Faculty of Psychology and Education Sciences – University of Porto (Portugal)

²University Institute of Health Sciences (CESPU) (Portugal)

Abstract

Over the past decade, increasing demands from national and European funding bodies for reliable and scientifically grounded evidence have reinforced the importance of Social Impact Measurement (SIM) in the evaluation of public policies. Within this context, SINCLab – Social Inclusion Laboratory has conducted several SIM studies applied to socio-educational innovation policies implemented at both local and regional levels in Portugal. This paper discusses the ethical, operational, and methodological challenges associated with the use of control groups in SIM studies conducted in real public policy contexts. While quasi-experimental designs are often considered the most appropriate approach for establishing causal inference in policy evaluation, the nature of socio-educational programmes, their universal implementation, and ethical constraints frequently limit the feasibility of control groups. Drawing on the evaluation of programmes such as the Integrated and Innovative Plan to Combat School Failure (PIICIE) and the Action Plan for Disadvantaged Communities (PACD), the paper discusses methodological alternatives that allow for robust impact inference. Ultimately, the paper argues that strengthening the methodological rigor of social impact measurement is essential for ensuring that socio-educational public policies are both scientifically grounded and capable of producing meaningful societal outcomes.

Keywords: *Social impact measurement, public policy evaluation, quasi-experimental designs, evidence-based policy, socio-educational inclusion.*

1. The problem

Over the last decade, SINCLab – Social Inclusion Laboratory has conducted an extensive programme of Social Impact Measurement (SIM) studies applied to projects promoted by Public Policy (PP) actors at both local and regional levels (Municipalities, Intermunicipal Communities, and the Porto Metropolitan Area), particularly within the framework of national socio-educational innovation policies.

Increasing demands from national and European funding bodies for reliable, independent, and scientifically grounded evidence regarding the outcomes and impacts of public policies have reinforced the importance of adopting evaluation methodologies capable of establishing empirically robust relationships between implemented interventions and observable changes in target populations.

Within this context arises the central issue addressed in this paper: how can researchers deal with the ethical, operational, and methodological challenges associated with the use of control groups in SIM studies conducted in real public policy settings?

Although the SIM models developed by SINCLab favour quasi-experimental designs—whenever possible using Control Groups (CG)—the nature of socio-educational policies, the population coverage of these programmes, and the ethical principles inherent to working with children and young people often make the establishment of such groups unfeasible. It therefore becomes necessary to discuss when it is possible, appropriate, or even desirable to dispense with CGs and to identify methodological alternatives capable of ensuring the scientific robustness of the evidence produced.

2. The conceptual approach

To frame the rationale underlying this discussion, the Conceptual Model of Social Impact and the Social Impact Measurement (SIM) Model proposed to two public policy actors were presented: the Tâmega e Sousa Intermunicipal Community (regional level) and the Municipality of Porto (local level). The work developed with these partners aimed to conduct Social Impact Measurement, in the first case, of the Integrated and Innovative Plan to Combat School Failure (PIICIE) implemented across the eleven municipalities of the intermunicipal community, and, in the second case, of an Action Plan for Disadvantaged Communities (PACD). Together, these programmes involved several tens of thousands of direct beneficiaries, mostly children and young people.

In simplified terms, and as part of promoting a culture of Evidence-Based Practices and Policies, a Conceptual Model of Social Impact of the public policy under analysis was defined in collaboration with these partners. As illustrated in the upper section of Figure 1, programmes such as PIICIE or PACD can be understood as public policy instruments through which it is possible to “control” or mitigate the effects of factors widely identified in the literature and public statistics as moderators of educational success and socio-educational inclusion (e.g., sociodemographic characteristics, territorial contexts, living conditions, and environmental contexts).

Although these moderating factors are addressed by a wide range of public policies, within socio-educational and school contexts the most appropriate policy “tool” to address them consists in promoting other factors that operate as mediators between the processes triggered by these moderators and the negative outcomes they may generate for educational success and socio-educational integration.

Increasingly, these factors—although not always explicitly labelled as *mediators*—are clearly identified in the strategic frameworks, objectives, and targets of public policies (for example, psychosocial competences, cognitions, attitudes, beliefs, expectations, motivation, and emotions). In other words, public policies in the domain discussed here increasingly focus on the promotion or activation of processes that positively contribute to educational success and socio-educational inclusion, as these processes may potentially mitigate or buffer the effects of the moderating factors mentioned above.

This strategic orientation becomes even more explicit within the broader framework of public policies designed to promote development, inclusion, and social cohesion. As illustrated by the examples presented in Figure 1, the vast majority of these policies advocate the promotion of mediating factors that are inherently psychological in nature and operate at multiple levels (e.g., intrapersonal, interpersonal, group, intra-group or inter-group processes).

If the central mechanisms targeted by these policies are indeed of this nature, it becomes evident that the social impact measurement of such public policies necessarily requires psychological knowledge and research, conducted by professionals with the appropriate expertise.

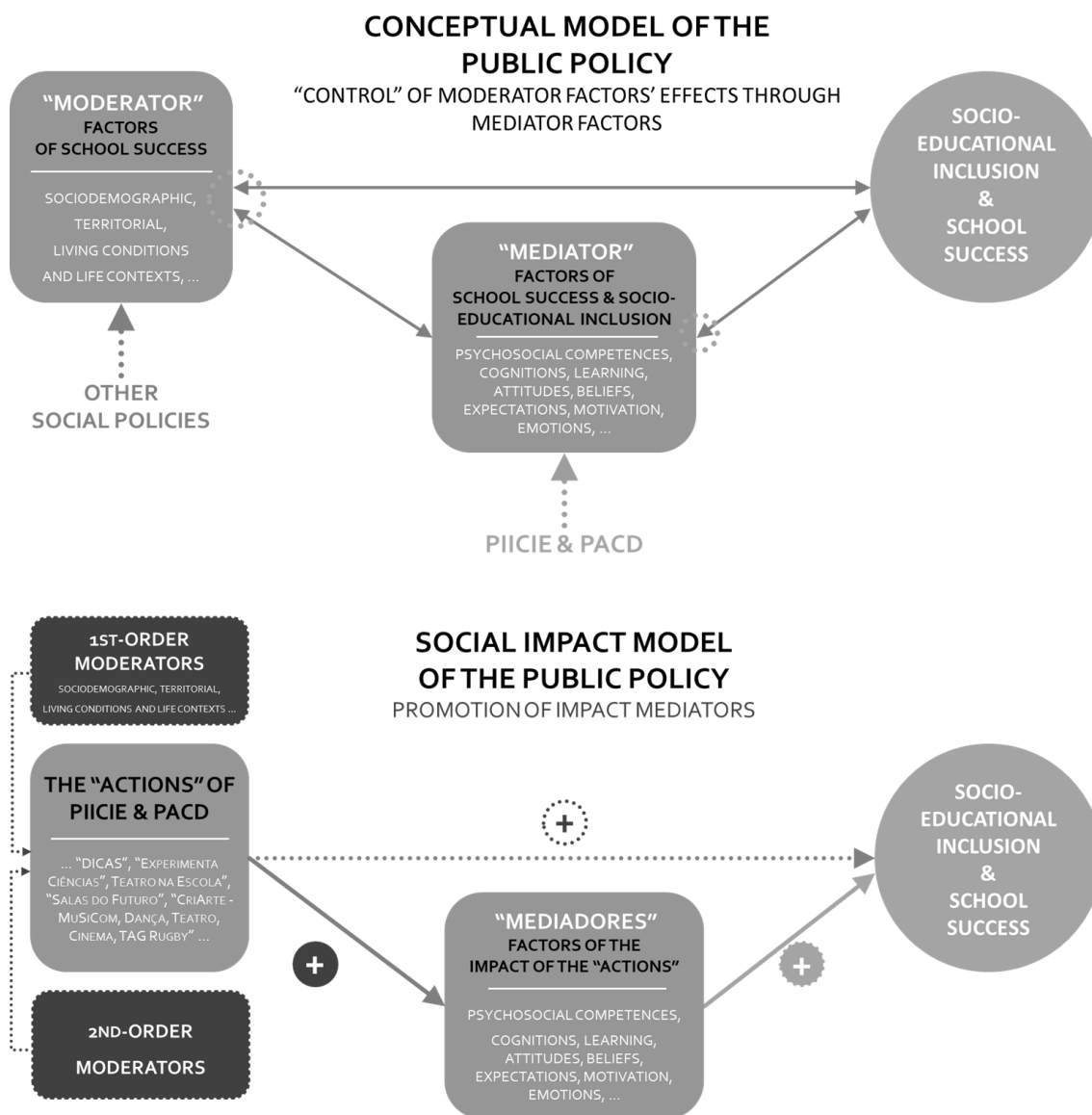
In the lower section of Figure 1, the Social Impact Measurement Model proposed to the two public entities mentioned above is illustrated. From a scientific perspective, three key assumptions can be highlighted:

- (1) the actions promoted by a public policy are not, in themselves, direct generators of the intended social impact;
- (2) rather, they function as instruments through which factors identified as positive correlates of the intended impact are promoted; and
- (3) from both a conceptual and statistical standpoint, it is precisely these mediating factors that can be considered Impact Mediators.

Therefore, the conceptual logic underlying this model assumes that public policy actions influence mediating psychosocial processes, which in turn contribute to socio-educational outcomes, while structural contextual factors operate as moderators of these relationships. As represented in Figure 1, the expected impact of public policies is thus produced through a mediated pathway between policy actions and outcomes, conditioned by contextual moderators.

In summary, if a public policy aims to promote these mediating factors, then the impact of that policy should be measured precisely at the level of those mediators.

Figure 1. Conceptual Model and Social Impact Measurement Model of a Socio-Educational Public Policy.



Note. The figure illustrates the conceptual relationship between moderators, mediators, and socio-educational outcomes within public policy interventions.

3. The methodologies and their results

3.1. “Discussing Methodology” to promote a culture and literacy of evidence-based public policies

The primacy of evidence has long been an established orientation among international organisations (United Nations Development Programme, 2009; UNEG Impact Evaluation Task Force, 2013; UNICEF Office of Research, 2014). However, only since the second half of the European 2020 policy framework has it gained stronger prominence within the context of national public policies (European Commission – GECES Sub-group on Impact Measurement, 2014). In practice, this orientation has increasingly evolved into a directive, particularly influencing local and regional public authorities.

Nevertheless, this “new culture” is rarely accompanied by an adequate level of scientific literacy among the actors involved. Put colloquially, it would be fair to say that the maxim “render unto Caesar what is Caesar’s” has not yet translated into practices grounded in the scientific principle that each methodology has its own merits and consequences.

For instance, both in discourse and in practice it is common to find the terms “impact inference” or even “impact attribution” being used without employing the methodologies that legitimately allow such claims — namely experimental or quasi-experimental approaches.

Within the partnership established with the two institutional actors mentioned above, and with other partners in additional projects, SINCLab (co)defined that the continuation, dissemination, or scaling-up of the socio-educational policies under analysis should be supported by evidence derived from methodologies that, based on well-established scientific principles, allow the testing of impact attribution, or, when this is not feasible, allow the testing of impact inference.

3.2. The Control Group “dilemma”

In the majority of projects in which SINCLab conducts applied research on the social impact measurement of public policies, the SIM models implemented rely on quasi-experimental approaches, most commonly pre- vs. post-test designs with Control Groups (CG). Whenever possible, Solomon designs are also employed, as they represent a more robust variant in terms of the external validity of the impact evidence collected.

However, as noted earlier, for ethical reasons (and many others beyond those previously mentioned) it is often impossible for SIM models to include control groups. Although the absence of CGs is frequently due to ethical and deontological imperatives associated with safeguarding the well-being of individuals, it is equally common for this to result from the fact that the public policy under evaluation is implemented universally across the target population.

In such cases, subject-wise randomisation — necessary to implement a randomised controlled trial (RCT) or similar model — becomes extremely difficult to sustain. When randomisation is feasible, a group-wise variant tends to be more appropriate, as it better reflects the collective implementation contexts typical of most public policies.

That said, we argue that the methodological and conceptual solution that best mitigates the constraints associated with the impossibility of including control groups is the introduction of factors within SIM models that can assume the statistical (and conceptual) status of moderators.

Although such models do not allow strict impact attribution, they nonetheless provide a scientifically robust basis for impact inference when such effects occur.

In summary, we argue that the evaluation of public policies should rely on quasi-experimental methodological approaches capable of producing evidence that demonstrates (or not) the existence of specific effects or impacts, because these effects occur only as a function of a clearly identified moderator, or because their magnitude varies depending on that moderator.

Put schematically: if a given intervention, action, project, programme, or public policy does not produce impact, then the moderator identified as associated with it will also fail to show an effect — for instance, a change between pre- and post-test measures. Statistically speaking, no interaction effect will be observed.

4. Conclusions: Solutions for maintaining standards of “evidence-based evaluation”

The issues discussed throughout this paper can be succinctly summarised by posing the following question: why “lower the standards” in the evaluation of public policies if sound methodological solutions exist to address the constraints of such research contexts?

Indeed, although it is well established that very significant differences exist between basic research and applied research, it is equally well established that the foundational principles of the scientific process constitute their common ground (Oskamp and Schultz, 1998; Steg et al., 2017). Consequently, those engaged in applied research on the evaluation of public policies should assume the responsibility of promoting the scientific literacy of public policy actors, so that these actors incorporate into their practices the commitment to grounding public policies in primarily scientific evidence, both in their design and in the measurement of their impact.

For example, within the so-called ecosystem of Social Innovation public policies, a wide range of methodologies have been employed that lack empirical and scientific grounding, and which nevertheless continue to be accepted as legitimate sources of evidence regarding their impact.

4.1. The “excuse” of costs and complexity

In the field of public policy evaluation it is common to encounter arguments based on the “excuse” that randomised controlled trials (RCTs) and quasi-experimental designs are expensive, but above all that they are so complex that stakeholders may not be able to incorporate them into their practices — or even understand them. However, there is no empirical evidence to support such claims.

To address the ethical, methodological, and logistical challenges that the real-world context of public policy evaluation inevitably entails, SINCLab argues that it is necessary to rely on methodological solutions grounded in sound scientific practices, such as classic or retrospective pre- vs. post-test designs, Solomon designs, time-series analysis, Stepped-Wedge models, or accelerated longitudinal designs.

If gold standards exist, then it is on these that the evaluation of public policies — particularly the evaluation of their impact — should be grounded. Moreover, if these policies explicitly define their central focus as the promotion of an increasingly broad range of psychosocial factors and processes, then psychology and its professionals (both researchers and practitioners) must also occupy a central role, both in their development and in their evaluation.

References

- European Commission – GECES Sub-group on Impact Measurement. (2014). Proposed approaches to social impact measurement in legislation and in practice relating to EuSEFs and the EaSI. European Commission.
- Oskamp, S., & Schultz, P. W. (1998). *Applied social psychology* (2nd ed.). Prentice Hall.
- Steg, L., Keizer, K., Buunk, A. P., & Rothengatter, T. (2017). *Applied social psychology: Understanding and managing social problems*. Cambridge University Press.
- UNEG Impact Evaluation Task Force. (2013). *Impact evaluation in UN agency evaluation systems: Guidance on selection, planning and management*. United Nations Evaluation Group.
- UNICEF Office of Research. (2014). *Methodological briefs: Impact evaluation*. UNICEF.
- United Nations Development Programme. (2009). *Handbook on planning, monitoring and evaluating for development results*. UNDP.