

A Critical Appraisal of Sure Start Evaluations

LIU Hai He Ye Li Min

Overseas Education College, Chengdu University, Chengdu, Sichuan, 610106, China

Keywords: Sure start, Policy evaluation, Early intervention programmers

Abstract: This essay has critically appraised policy evaluation in Sure Start initiative in the UK. Sure Start Local Programmes (SSLPs), initiated by UK Government, aimed at addressing child poverty and social exclusion. The impetus for the implementation of Sure Start was built on the previous research, particularly evidence from the American early intervention programmes. This study focused on one of the evidence, a longitudinal study on parent involvement. By examining the design and rigour of the study, it is concluded that the use of one-group pre-test and post-test design cannot address the effectiveness of the intervention.

1. Introduction

In July 1998, a series of Sure Start Local Programmes (SSLPs) has been established by Gordon Brown, which aimed at tackling child poverty and social exclusion and targeted at children under 4 years old and their families who live in areas of high deprivation (Clarke, 2006). And the purpose of the intervention is to enhance these children's physical, intellectual, social and emotional status, enabling them to realize their potential within the education system, so as to avoid educational problems, such as juvenile crime, unemployment and teenage pregnancy (Robert, 2000).

In October 1997, a cross-departmental Review of service for young children was set up. The Review was to examine current services for children aged seven and under, which focused on social exclusion and young children by means of looking at policies and services devoted to children, families and communities and ensuring their effectiveness to 'ensure the development of their full potential throughout their lives' (Glass, 1999, p. 260). According to the Review, there are some key elements considered to the background of this initiative. Firstly, the level of poverty in the UK was very high on the basis of European standards and disadvantage among young children was increasing. Secondly, the current services for young children were highly varied, uncoordinated and patchy (Glass, 1999). Furthermore, the Review also found that children under four were missed out from the services provision. Therefore, it is recommended that there should be a change in service design and delivery.

2. Evidence on Which Policy Was Based

As mentioned before, the Sure Start programme emerged from a Cross-Departmental Review of Provision for Young Children (H.M., Treasury 1998). While there were many influences on the Review (Welshman, 2010), as Glass (1999) argues, the Review was heavily influenced by a number of programmes in the United States, which showed the evidence for the effectiveness of early interventions. Among these programmes, evidence that were influential included Perry Pre-School / HighScope Project and both short- and long- term evaluations of the Head Start, where longitudinal research have demonstrated that provision of high quality childcare services for children from poor families, combined with home visiting, producing worthwhile benefits for children, families and communities (Belsky, Mulhish & Barnes, 2008).

However, hundreds of studies have evaluated the Head Start policy and they are varied widely in subject, design, topics and findings. Among which, some studies were selected to justify the Sure Start programme, while some were not used in the policy-making process. Since the findings of Head Start studies differ widely, in this study, a longitudinal study on parent involvement was critically appraised through looking at its design and rigour (Parker et al., 1997).

Head Start was a two-generation program aimed at meeting the needs of both parents and their children. Although parents were involved in all aspects of the intervention, many studies have only focused on its effectiveness on children's

cognitive development and educational attainment. Ignoring the potential benefits of parent participation for child and families may underestimates the full benefits of Head Start (Parker et al., 1997). Therefore, as Zigler (1978) suggested, evaluation of Head Start should pay more attention on to parents. This longitudinal study addressed the issue and tried to evaluate the effectiveness of Head Start in two ways: firstly, to assess the effects of parents' participation on the parent, their child and an older sibling; secondly, to examine the demographic, contextual and family characteristics which related to the extent of parents' participation in the program (Parker et al., 1997).

By means of a quasi-experimental pretest posttest design, this study has a lot of strengths and limitations. In terms of strengths, firstly, compared to previous evaluation on Head Start, the pretest data collected could control for preexisting differences among the parents in the outcome and important demographic differences (e.g., mother's age, education). Secondly, four types of parental activities assessed in the study which helped researchers to better understand the complexity of parents' experiences. Thirdly, as this study followed the children into elementary schools, therefore, longer time effects of Head Start can be assessed. Fourthly, in this study, there is no controlled group of non-Head Start parents and children, therefore, researchers were allowed to examine the effects of naturally-occurring variations in parent involvement (Parker et al., 1997).

With respect to the limitations, the first problem with a pre- and post- test design is one of temporal trends. In this study, the positive outcomes for the child may irrespective of the intervention, but simple through increased maturity during the passage of time. Secondly, since the implementation of Job Opportunities and Basic Skills (JOBS) welfare-to work policy largely affected the Head Start programs, the Cohort I and Cohort II became non-comparable on the major variable, it is unfeasible to make combine cohorts. Therefore, without a comparison group, this study cannot control for regression to the mean and temporal effects (David, 2001). Thirdly, this study cannot make an inference for causality because it is possible that personal, demographic and contextual factors affect the original pretest scores, which would lead to differences in outcomes.

Overall, this study showed that parent involvement in Head Start benefits parents and children. However, since there are a lot of limitations about the design, the validity of outcomes was largely affected.

3. Evaluations

Three key evaluations will be discussed in this study. The first one was a pragmatic randomised controlled trial – “parenting intervention in Sure Start services for children at risk of developing conduct disorder” (Hutchings et al., 2007). The background of this study was discussed as followed. Firstly, there is an increasingly number of young people in the UK conduct disorder and the percentage is around 5%-10%. Furthermore, it is found that the situation was particularly serious in areas of disadvantage (Belsky, ect. 2006). It is showed that children who have behavior problems in their early age are more likely to conduct antisocial and criminal behavior in their adolescence (Broidy et al., 2003). As Webster-Stratton (1998) analyzed, one of the reasons for children’s problem behaviors is that their parents lack key parenting skills. However, in the UK, early parenting support programme of Sure Start (Hutchings, ect., 2004) showed no significant effect in reducing children’s’ behavior difficulties. Therefore, built on previous experiences, the Webster-Stratton Incredible Years basic parenting programme took all potential factors which were regarded improving parent training outcomes into consideration, used 11 Sure Start services in north and mid-Wales since 2001.

On the other hand, this study has a number of weaknesses as well. Firstly, this research only set in north and mid-Wales, the results are not necessarily applicable to other area of UK, which were not externally valid and cannot generalized beyond the Wales. Secondly, the sample size is small. In this study, 12 groups in 11 areas in total of 153 families are evaluated, which affect the internal validity of this study and cannot detect important differences (Torgerson & Torgerson, 2008). Besides, this study had six months follow-up, which was short, and reduce the strength of the intervention.

The second and third studies were from The National Evaluation of Sure Start (NESS). Commissioned in 2001, the NESS conducted by a group led by Professor Edward Mulhish (Melhish, ect., 2010). It is made up of five sections: Implementation, Impact, Local Context Analysis, Cost Effectiveness and Support to Local Programmes. Funded by the Sure Start Unit, the evaluation studied the effectiveness of Sure Start programmes in England (NESS, 2005). There are three purposes of the implementation evaluation: firstly, to provide data on key aspects of design, policy, practice, style and development in each local Sure Start programmes; Secondly, to produce qualitative information on services as a framework for the impact study; Thirdly, to provide data on programme objectives and facilitate the quantification of inputs for the analysis of cost effectiveness.

In this study, two evaluations were chosen from the NESS, the earliest one published in 2005 and the latest one published in 2012. In 2005, the NESS conducted an evaluation named “Early findings on the impact of sure start local programmes on child development and family functioning: report of the cross-sectional study of 9-and 36- Month old children and their families” (NESS, 2005). This evaluation summarized preliminary reports in which data collected from the cross-sectional study and presented findings evaluate effectiveness of SSLPs in enhancing the well being of 9-month olds (who will be studied again at ages 3 and 5) and 36-month olds (who are only studied once) and their families. This study compared the SSLP areas and Sure Start-to-be areas, which recruit 12000 9-month-olds and 3000 36-month-olds and their families from the 150 SSLP areas, and 1250 families with 9-month olds and 1250 families with 36-month olds from 50 comparison communities by the end of 2004 (NESS, 2005).

Using the cross-sectional design, the evaluation has a lot of strengths and limitations. In terms of advantages, firstly, since the data are collected at one point of time, this study avoid the threats like history, maturation, instrument decay, statistical regression, mortality and testing effects (David, 2001), which arise from the experimental design. In addition, data are obtained relatively quickly by using the cross-sectional design, because researchers do not need to wait for various follow-up stages or

interventions before analysing the data (David, 2001). Secondly, data are collected from a variety of ways, for example, parental report, observation and developmental assessments. The use of methodological triangulation technique could facilitate the validation of the data (Bogdan & Biklen, 2006). Thirdly, in the data analysis stage, in order to tackle confounding variables, this study collected demographic and background information from each family, as well as area characteristics on each community. These data was thought to potentially influence the outcome measures and to differ between SSLP and Sure Start-to-be communities, which could be taken into account in the statistical analysis of the data.

However, there are still some drawbacks. Firstly, in the cross-sectional design, 'groups' are constructed on the basis of existing differences in the sample. And according to the different category of the independent variable, participants are divided up into groups. In this study, two comparisons groups are constructed, SSLP areas and areas were have such programmes shortly after data collection (i.e. Sure Start-to-be communities). Therefore, this makes this study difficult to work out whether any group differences on the outcome available are due to the programmes or to other related uncontrolled differences between the groups. Secondly, compare to experimental design, cross-sectional study can only controls for variables that they have thought of and about which they have information, they can never be sure that they controlled for all relevant variables (David, 2001). In this study, although some variables, like age, gender, maternal work status and so on were considered in the data analysis stage, researchers cannot make sure that they have controlled for all relevant variables, which influence the internal validity of this study.

In 2012, in order to assess the impact of SSLPs on child and family functioning over time, the National Evaluation of Sure Start (NESS) has followed up appropriately 5000 7 year-olds and their families in 150 SSLP areas who were initially studied when the children were 9 months, 3 and 5 years old. And the evaluation aimed to investigate whether differences in child and family functioning found at 3 and 5 years of age persist until 7 years of age, and whether any other differences emerge. Two groups are identified to compare in this study: one from SSLP areas and another from areas where did not offer SSLP service, but children in both these areas share the similar characteristics (NESS, 2012).

There are a number of methodological and practical issues in the longitudinal study. Firstly, as for the research design, it is widely acknowledged that Randomized Controlled Trials (RCTs) are considered the 'gold standard' for estimating the effects of treatments, interventions, and exposures on outcomes (Torgerson & Torgerson, 2008). In this study, since Sure Start programmes targeted at areas rather than individuals, and it expanded rapidly, it is not appropriate to use RCTs. Without the randomization, selection bias would occur. In order to eliminate the selection bias, this study used the Propensity Score Analysis (Rosenbaum & Rubin, 1983), which estimate the likelihood of being a SSLP area by distinguishing between groups on area characteristics. In addition, the NESS team adopted an approach based upon quasi-experimental methods, which was regarded the next best evaluation design. This strategy allowed for effects of child, family and community characteristics, which answered the question of whether SSLPs have an effect (NESS, 2012). However, although the evaluation statistically controls for many relevant covariates, some unmeasured and unknown differences (e.g. genetic factors) may be responsible for the outcomes.

Secondly, with respect to the data collection, data collected two years apart by two different research teams in this study. Therefore, the effect of SSLPs could be attributed to other factors, for example, history and maturation, as this study extends over time. Besides, although close cooperation between teams and cooperation in staff training are implemented, the effect of SSLPs could also be influenced as the result of differences in measurement by the two research teams. Furthermore, it is unavoidable

that data are missed in this study, for example, some family could not be contacted or some funders do not decide to follow up all those seen at 5 years of age when they were age 7. These missing data could introduce possible bias to the outcomes. However, in order to counter these possible bias, researchers used the group means approach, which divide the sample into three different but overlapping samples, and obtained the mean for the missing data variable for each of these subgroups (David, 2001). This approach through using all the other available information on all individuals to calculate a missing value, and then taken into consideration the likelihood of error in such estimates to reduce the bias (NESS, 2012).

4. Critical Issues

There are some debates among the Sure Start programmes. According to Rutter (2006), firstly, he criticized that the selection of Sure Start areas led to some problems. Since the designated areas were chosen on the basis of the level of deprivation, some seriously disadvantaged families lived out of these areas would be excluded. Secondly, he argued, as SSLPs are area-based programmes, which lack a prescriptive prescribed curriculum or services (Rutter, 2006, p. 135). Instead, SSLPs allowed each area to create and improve services as needed. In this case, SSLPs are highly varied, which result in difficulty of evaluation. However, Davies (2007) disagreed Rutter's opinion and argued that based on clear targets, the SSLPs do have a clear curriculum. And the advantage of SSLPS was that each area was allowed to take actions which were appropriate. But Rutter (2007) refuted, evidence from the relevant documents showed that, SSLPs specified what the targets should be in each area, but they do not explained clearly how to achieve these goals. Thirdly, Rutter (2006) analysed the reasons why the Government rule out any form of Randomised Controlled Trials (RCT) design, though it was regarded the most effective method for addressing effectiveness (Torgerson & Torgerson, 2008). He concluded that RCTs are only used when the effectiveness of an intervention is unsure. As for the Sure Start, the Government believed it would drastically reduce child poverty and social exclusion. As some cynics pointed out, if RCT showed that the Sure Start was ineffective, which meant that the policy implemented by Government was wrong (Rutter, 2006). Fourthly, there is a debate about the successes of Sure Start. According to Davies (2007), there are three great successes of Sure Start: great local involvement, holistic services and increased participants. Rutter (2007) agreed his point to some extent, but he added that, although SSLPs had many useful initiatives and made great positive effects on children and families, some ineffective or even counter-productive elements were also identified. He also suggested that a randomised controlled trial or some kind of quasi-experimental designs which could really provided evidence on the effectiveness of Sure Start should be adopted.

5. Conclusion

This paper has discussed the advantages and limitations of three key evaluations on Sure Start . According to Torgerson & Torgerson (2008), a randomized controlled trial is the best way to address causality, as it eliminates selection bias, controls for regression to the mean effects, temporal effects and also provides a basis for statistical inference. Therefore, by means of the RCT, the 2007 study of parenting intervention within the Sure Start systems in Wales addressed the effectiveness of Sure Start. However, the NESS of 2005 and 2012 cannot match an impact question of Sure Start because there are several limitations within the cross-sectional and longitudinal design. Through comparing these three

evaluations, it is evident that each design has its own advantages and limitations, but for the effectiveness questions, it is best to use Randomized Controlled Trials.

6. Acknowledgment

A Project Supported by Center for Early Childhood Education Research, Sichuan. Project name is Research on Kindergarten Art Field Curriculum under the Vision of Life Aesthetics, Project No. is CECER-2020-B05.

References

- [1] Belsky, J., Mulhish, E., Barnes, J., Leyland, A. H., & Romaniuk, H. (2006). *Effects of Sure Start local programmes on children and families: early findings from a quasi-experimental, cross sectional study*. *BMJ*, 332(7556): pp.1476-1478.
- [2] Bogdan, R.C. & Bicen, S. K. (2006). *Qualitative research in education: An introduction to theory and methods*. Allyn & Bacon.
- [3] Broidy, L.M., Tremblay, R. E., Brame, B., Fergusson, D., Horwood, J. L., Laird, R., ... & Vitaro, F. (2003). *Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study*. *Developmental psychology*, 39(2): p.222.
- [4] Clarke, K. (2006). *Childhood, parenting and early intervention: A critical examination of the Sure Start national programme*. *Critical Social Policy*, 26(4): pp.699-721.
- [5] Davies, J. (2007). *Is Sure Start an Effective Preventive Intervention?*. *Child and Adolescent Mental Health*, 12(1): pp.55-55.
- [6] De Vaus, D. (2001). *Research design in social research*. SAGE Publications Limited.
- [7] Eisenstadt, N. (2002). *Sure Start: key principles and ethos*. *Child: Care, Health and Development*, 28(1): p.3-4.
- [8] Glass, N. (1999). *Sure Start: the development of an early intervention programme for young children in the United Kingdom*. *Children & Society*, 13(4): pp.257-264.
- [9] Hutchings, J., Gardner, F., & Lane, E. (2004). *Making evidence-based interventions work. Support from the start: Working with young children and their families to reduce the risks of crime and anti-social behaviour*, pp.69-79.
- [10] Hutchings, J., Bywater, T., Daley, D., Gardner, F., Whitaker, C., Jones, K., ... & Edwards, R. T. (2007). *Parenting intervention in Sure Start services for children at risk of developing conduct disorder: pragmatic randomised controlled trial*. *BMJ*, 334(7595):p. 678
- [11] Hollis, S., & Campbell, F. (1999). *What is meant by intention to treat analysis? Survey of published randomised controlled trials*. *BMJ*, 319 (7211): pp.670-674.
- [12] Melhuish, E., Belsky, J. & Barnes, J. (2010). *Evaluation and value of Sure Start*. *Archives of disease in childhood*, 95(3): pp.159-161.