Variation in community intervention programmes and consequences for children and families: the example of Sure Start Local Programmes

Edward Melhuish,1 Jay Belsky,1 Angela Anning,1 Mog Ball,1 Jacqueline Barnes,1 Helena Romaniuk,1 Alastair Leyland,2 and the NESS Research Team1

1Institute for the Study of Children Families and Social Issues, Birkbeck, University of London, UK; 2MRC Social & Public Health Sciences Unit, University of Glasgow, UK

Background: An area-based initiative, Sure Start Local Programmes (SSLPs), was established by the UK government to reduce social exclusion through improving the well-being of children aged 0–3 years and their families in disadvantaged communities; a true community intervention in that all children under four and their families in specified areas served as targets of universal services. A national evaluation examined the links between variation in programme implementation and effectiveness.

Methods: Data gathered from multiple sources produced measures of implementation in terms of proficiency, services and staffing. Measures of programme impact on child/parenting outcomes derived from multilevel models, controlling for child, family and area characteristics, were identified to demonstrate programme effectiveness. Results: Some modest linkage between programme implementation (e.g., proficiency, empowerment of parents and staff, identification of users) and effectiveness for child and parenting outcomes. Conclusions: Overall proficiency and specific aspects of implementation may influence effectiveness, which should guide the design of other child, family and community services. Keywords: Child development, parenting, community interventions, programme implementation, programme effectiveness, Sure Start. Abbreviations: SSLP: Sure Start Local Programme; NESS: National Evaluation of Sure Start; CCI: Comprehensive Community Initiative.

In the 1980s and 90s several Comprehensive Community Initiatives (CCIs) emerged in the United States, founded on two principles: the need for strong partnerships between families, governments, child welfare, family support, health, educational agencies and other organisations; and the need to empower community members to participate actively with professionals, to promote healthier communities (Tomison & Wise, 1999). Most initiatives also emphasised devolution of authority from higher levels to the community (bottom-up) in an effort to enhance child/parent well-being (Kubisch, Weiss, Schoor, & Connell, 1995).

CCIs targeted a range of outcomes – teen pregnancy, youth employment, and crime, as did more broad-based interventions such as Empowerment Zones and Enterprise Communities (Kubisch et al., 1995). One programme, Neighbors Helping Neighbors: A New National Strategy for the Protection of Children, addressed parenting and specifically child maltreatment through reorienting services to the neighbourhood level (US Advisory Board on Child Abuse and Neglect, 1993). Rather than suggesting particular strategies, the emphasis was on promoting a culture of responsibility and planning in communities. In 1989, the US National Center on Child Abuse and Neglect (NCCAN) funded nine model comprehensive community-based projects, with each community shaping its own services. A national evaluation found that local task forces were central to success in implementing these initiatives (CSR, 1996).

In the UK several ‘Area Based Initiatives’ (ABIs) have emerged since the late 1990s. They have broader perspectives than those in the USA, with a focus on narrowing the gap between deprived and non-deprived neighbourhoods as a means of tackling social exclusion and improving services. While the majority concern economic aspects of communities, some target children and parents. As a cornerstone of the UK government’s campaign to reduce the intergenerational transmission of child poverty and social exclusion, Sure Start Local Programmes (SSLPs) were established in 1999 (Glass, 1999). SSLPs were to improve the well-being of children aged 0–3 years and their families living in disadvantaged neighbourhoods so children would flourish later in life. SSLPs were to be part of a seamless range of support for children at risk of social exclusion, linking up with other government initiatives, including the Children’s Fund, providing services for children aged 5–13 years, and Connexions, for ages 13–19 years.

Each SSLP was to be a comprehensive, community-based programme adapted to local needs, using local expertise, and capitalising on shared concerns of
community residents (Oliver, Smith, & Barker, 1998). By targeting disadvantaged communities rather than at-risk individuals they were to be stigma-free. SSLPs were charged with improving children’s social/emotional development, health and learning, and strengthening families and communities. There were 260 SSLPs by 2001, and 524 SSLPs existed by 2004. Such rapid expansion limited the opportunity for later programmes to learn from earlier ones.

SSLPs were true community interventions (see Barnes, Katz, Korbin, & O’Brien, 2006), with all children under four and their families in an area serving as the ‘targets’ of intervention. Community control was emphasised and was to be exercised through local partnership management boards, to bring together everyone in the community concerned with children (i.e., health services, social services, education, private sector, voluntary sector and parents). To facilitate this bottom-up approach there was no specification of how to provide services, only what they should achieve. However, all SSLPs were expected to provide five core services: 1) outreach and home visiting, 2) support for families and parents, 3) good quality play, learning and childcare, 4) primary and community health care and advice about child and family health, and 5) support for children and parents with special needs.

Programmes aimed to improve existing services and create new ones as needed, without specification of how services were to be changed or what exactly was to be delivered. This contrasts markedly with more focused early interventions demonstrated to be effective, be they childcare based, like the Abeer-darian Project (Ramey et al., 2000); home based, like the Prenatal Early Intervention Project (Olds et al., 1999) or the Positive Parenting Program (Sanders, 2003); or even a combination of centre and home based, like Early Head Start (Love et al., 2005).

While all communities selected were disadvantaged, it has been established that disadvantaged areas, such as those selected for SSLPs, differ in important ways that lead to differences in child and family well-being (Barnes et al., 2005). Also, the local autonomy, together with the diverse history and service provision of the communities where SSLPs were established, and diversity among lead agencies and professionals, resulted in wide variation across the 524 programmes in what they do, how they do it, and in their proficiency in addressing objectives, posing a challenge for evaluation.

The National Evaluation of Sure Start (NESS) focused on the first 260 SSLPs and has detected some limited but significant across-the-board effects of SSLPs upon family functioning, after controlling for multiple child, family and area characteristics (Belsky et al., 2006). Mothers of 9-month-olds living in SSLP areas reported less household chaos and mothers of 36-month-olds showed greater parental acceptance relative to families living in similar areas that were scheduled to become SSLPs but had not yet done so. At the same time, it was clear that great variation existed in the degree to which various programmes benefited children and families.

**The current study**

Variation in intervention effectiveness has frequently been studied for individual-level highly specified interventions such as parent training (e.g., Hartman, Stage, & Webster-Stratton, 2003; Kazdin, Holland, Crowley, & Breton, 1997). Such work is based on the premise that even generally effective interventions vary in their impact and that illuminating the processes by which they prove to be more and less effective is critical for understanding not only how they achieve – or why they fail to achieve – their stated goals, but how their effectiveness can be increased. However well established this strategy of inquiry is within narrowly focused intervention programmes, little such research has been undertaken in the case of community-level interventions. Thus, having discerned general intervention effects of SSLPs using a ‘between-group’ design comparing families residing in communities with and without an SSLP (Belsky et al., 2006), the ‘within-group’ difference question arose: ‘Does variation in programmes account for variation in their impact on child/family functioning?’ By seeking to elucidate features of programmes that may partly account for why some might have more beneficial impacts on children and families than others, the current study is similar to Love et al.’s (2005) evaluation of variation in Early Head Start implementation.

Fidelity is the degree to which an intervention adheres to the originally developed protocol or programme. In evaluation research, the fidelity of implementation to the model of intervention is considered central to a programme’s success (e.g., Borrelli et al., 2005). Fidelity criteria may include programmatic structure, the framework for service delivery, and the ways services are delivered. Such criteria facilitate replication of an intervention (Bond et al., 2000) and, with highly specified programmes, fidelity is central to the study of programme variation. In the case of SSLPs, however, the degree of local autonomy, combined with the lack of specification of how the programmes’ aims were to be achieved, meant that there was no specified model against which fidelity could be formally evaluated.

To address this challenge, the evaluation developed an alternative to programme fidelity, namely, proficiency in realising the principles that were the basis for SSLPs, as defined in central government guidance documents (Sure Start Unit, 2002). Scales to rate the 18 identified principles (see Table 1) were developed to assess programme quality and, thereby, likelihood of benefiting children/families (Zigler & Styfco, 2004). Collectively, these principles reflect the implicit and explicit ‘theory of change’ guiding SSLPs (see Weiss, 1995). For example, because
Table 1 Eighteen domains of SSLP implementation proficiency

<table>
<thead>
<tr>
<th>Process (N = 7)</th>
<th>Progress (N = 7)</th>
<th>Holistic (N = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partnership – composition: SSLP Partnership Board has balanced representation of education, social services, health, voluntary and community organisations and parents.</td>
<td>Services – quantity: Service delivery reflects guidance for core services in family support, health, play, early learning and childcare.</td>
<td>Vision: SSLP has a well-articulated vision relevant to the community.</td>
</tr>
<tr>
<td>Partnership – functioning: The Partnership functions well. Leadership: SSLP has effective leadership/management. Multi-agency working: Multi-agency teamwork is well established.</td>
<td>Services – delivery: SSLP has balanced focus on children, family and community. Identification of users: SSLP has strategies for identification of users. Reach: SSLP shows realistic and substantial involvement of families. Reach strategies: SSLP has strategies to improve and sustain use of services.</td>
<td>Empowerment: SSLP procedures create an environment empowering users and staff.</td>
</tr>
<tr>
<td>Service access: There are clear pathways to access specialist services. Staff turnover: Staff turnover low. Evaluation use: SSLP takes account of evaluation findings.</td>
<td>Services – innovation: SSLP shows innovation in service delivery. Services – flexibility: Services accommodate the needs of a wide range of users.</td>
<td>Communications: Communications reflect characteristics/languages of community.</td>
</tr>
<tr>
<td>Benefits to children and families were expected to derive from successful efforts to build functional partnerships among stakeholders, to empower service providers and parents alike, and to facilitate family access to available services, each of these domains of SSLP implementation was measured.</td>
<td>Holistic (N = 4)</td>
<td></td>
</tr>
</tbody>
</table>

Methods

Sample

One hundred and fifty of the first 260 SSLP areas were randomly sampled, stratified across the nine Government Office regions within England (see Belsky et al., 2006; NESS, 2005), to investigate programme impact on children/families. Families with 9- and 36-month-olds in these areas were randomly selected for recruitment using national Child Benefit records during 2003 and 2004. Home-visit data were gathered on 12,575 9-month-olds and 3,927 36-month-olds, representing response rates of 84.4% (9-month) and 73.4% (36-month). Mother/guardians provided written informed consent and a Multi-Regional Ethics Committee approved the research. In this report the focus is upon the 150 SSLPs that serve as units of analysis for this ‘within-group’ study.

Procedures and measures

SSLP implementation. Producing measures applicable across diverse programmes, which purportedly had the same aims, required innovative methods. Building on a pilot study, extensive quantitative and qualitative data from a range of sources were systematically collated, analysed and synthesised, thereby affording reliable quantitative ratings on domains of implementation proficiency. In addition, measures of separate types of service provision were produced.

Scales for rating the proficiency of implementation. Eighteen domains of implementation proficiency (7 about process, 7 about progress and 4 holistic aspects of implementation; see Table 1) were rated based on diverse data assembled on each SSLP. Data sources included programme delivery plans, an extensive NESS survey of SSLPs’ policies and practices supplemented by telephone follow-up; case study data where available; programme publications, publicity materials and organisational diagrams; SSLP local evaluation reports (see http://www.NESS.bbk.ac.uk); NESS staff appraisals of local evaluations; and SSLP expenditure and monitoring data (on numbers of families using programmes) provided by the government. Further data came from structured telephone interviews with key informants with insight into the histories, implementation and proficiency of the SSLPs, including: Programme Development Officers who supported SSLPs regionally; chairs of SSLP Management Boards; Local Authority Early Years Officers and NESS staff supporting local evaluations.

For each of the 18 domains a 7-point rating scale (1 = inadequate, 7 = excellent) was used with explicit criteria specified for each scale point (see Appendix 1 in supplemental materials on publisher’s website). Table 1 provides illustrative statements of the excellent end of each scale. Following training, one team of data collators searched the information gathered on each of the 150 SSLPs, separately assembling information specifically pertinent to each of the 18 ratings. Then, two trained raters, working independently, rated each programme on each of the 18 proficiency dimensions after reviewing the relevant assembled material. The level of inter-rater agreement within one point averaged 87% (range 77% to 98%) between these two raters, with weighted Kappa averaging .77 (range .55 to .97). For all 18 ratings the full range (1 to 7) was used and the distributions across the 150 programmes approximated normal distributions. The peak of the distribution was usually 3 or 4 (moderate scores) and the means varied from 3.65 to 4.32 with standard deviations varying from 1.01 to 1.51.

With regard to the construct validity of these 18 ratings, it is important to realise that ratings arise from inspection of accumulated data from diverse sources relevant to each dimension. The scales are anchored by criteria for every point on the 1–7 scale. The fact that independent raters apply similar ratings to a programme from inspecting independently collected data and applying the criteria is equivalent to showing that the ratings relate to the independently collected data on programme characteristics, i.e., that the ratings have face validity. One way to look at the predictive or
criterion validity of these scales is to consider their relationships with outcomes for the users of programmes. This is the focus of this report.

**Services and staffing.** Information concerning service provision was collated into four categories of service: 1) child-focused (e.g., early education/care, outside play areas, and 'BookStart'); 2) parent-focused (e.g., help-lines, health promotion, respite care, drop-in créches); 3) family-focused (e.g., family support, health services, family planning, and toy libraries); and 4) community-focused services that were available to all (e.g., welfare rights advice, credit unions, GP surgeries, and self-help groups). These four categories were subcategorised as inherited (predating SSLPs), improved (by SSLPs) or new (SSLP-created) services and the number of different services offered within each subcategory was counted. Staffing numbers in terms of full-time equivalents were converted to proportions of SSLP staff (adjustment for SSLP size) engaged in the four services, present in all programmes, of outreach, family support, health services, and play and childcare.

**Child and family outcomes.** During a 90-minute home visit, the mother/guardian was interviewed, and four subscales of the British Ability Scales (BAS II: Eliot, Smith, & McCulloch, 1996) were administered to English-speaking 36-month-olds: verbal comprehension, naming vocabulary, picture similarities and block-building. Although numerous child and parenting outcomes were measured (see Belsky et al., 2006), many were composited on the basis of factor analysis, and only those eight showing significant community-level variation in multilevel models qualified for inclusion in this study (NESS 2005). These included two 9-month parenting measures: maternal acceptance (i.e., avoidance of scolding/spanking/restraining) from the HOME Inventory (Caldwell & Bradley, 1984), and household chaos (i.e., extent that household is disorganised, noisy, lacking regular routine) (Matheny, Wachs, Ludwig, & Phillips, 1995); three 36-month parenting outcomes: maternal acceptance, negative parenting (a composite of reported parent/child conflict and closeness (Pianta, 2001), harsh discipline (e.g., swearing, threatening, smacking; Straus, Hamby, Finkelhor, Moore, & Runyan, 1998) and household chaos, and home learning environment (reflecting learning activities in the home, e.g., reading, learning songs, playing with numbers; Melhuish, Sylva, Sammons, Siraj-Blatchford, & Taggart, 2001); and three child outcomes for 36-month-olds: verbal and nonverbal ability from the British Ability Scales, and social competence (composite of ‘pro-social behaviour’ from the Strengths and Difficulties questionnaire, Goodman, 1997; and ‘independence’ from the Child Social Behaviour Questionnaire, Sammons et al., 2003). These child and parenting variables are used as outcomes in multilevel models to produce measures of SSLP effectiveness as described later in the preliminary analyses of the results.

**Family/community covariates.** Family-level covariates included child age, gender and ethnicity; maternal age, education, employment status, occupational status, and cognitive difficulties; father presence and employment; English language usage (primary, secondary, no English); and household income. Population characteristics of each SSLP area were obtained from diverse sources (e.g., census, DWP benefit records, school-achievement records, see Barnes et al., 2005) yielding, after factor analysis, a series of community-level covariates reflecting high vs. low prevalence of characteristics indicative of 1) Indian-subcontinent populations and proportion of young children, 2) Black (African or Caribbean) population and working-age adults, 3) lone and teen parents, 4) unemployed adults, 5) ill and disabled children, 6) infant mortality, 7) age 7 school achievement, 8) household crowding, 9) council housing and 10) adults in poor health.

**Analysis**

The analyses proceed from the more global to more micro-level. Initially multilevel modelling was used to derive measures of effectiveness for each programme. Then the ability of the 18 implementation ratings collectively to differentiate between more and less effective programmes was evaluated (via discriminant function analysis). Thereafter, regression analyses explored the ability of individual measures of implementation proficiency, services or staffing to predict programme effectiveness with respect to individual child and parenting outcomes.

Two sets of multilevel analyses were carried out that created measures of programme effectiveness, firstly using complete cases with no missing data and secondly using multiply imputed data where missing values were estimated using multiple-imputation procedures (Schafer, 1999). Using the first set of analyses with only complete cases avoids assumptions about missing data, but may lead to bias, as families with missing data may be different. Using the second set of analyses with imputed data avoids bias, but retains uncertainty about the basis for imputation. In this report the conservative strategy of using only results significant for both sets of analyses, i.e., complete-case and imputed data, was adopted.

**Preliminary analyses of programme effectiveness**

Children and families were nested within communities and hence the data are hierarchical. Potentially there is greater similarity between participants within the same communities so the independence of measurement assumption of regression is violated and mis-estimation of levels of significance is likely. Multilevel modelling overcomes such problems, taking into account the hierarchical data structure and producing more accurate predictions, and estimates of differences between participants, between communities, and between other comparison groups. Hence multi-level modelling was used, which also provides estimates of community effects that are equivalent to community ‘effectiveness’ scores for each community for each outcome. Indeed, these reflect the extent SSLPs produced outcomes better (or worse) than expected given child, family and community background characteristics. After controlling for child, family and area characteristics,
unexplained variation was partitioned into that attributable to the community and that attributable to the individual. Where significant community-level variation exists, as it does for the eight outcomes, the community-level standardised residuals become community-effectiveness scores, serving as the outcomes to be explained in this study treating the 150 communities as the units of analysis.

**Results**

The analyses address four specific questions.

**Do implementation ratings collectively discriminate programmes varying in effectiveness?**

To address this question, discriminant function analyses (Huberty, 1984) were undertaken. Specifically, standardised community residuals were averaged within a programme, separately for 9- and 36-month outcomes, and the distributions of the resulting composites of (9- and 36-month) effectiveness scores were divided at the median to create groups of more and less effective programmes. The results of these analyses are shown in Table 2. Collectively, the 18 programme-proficiency ratings significantly discriminated between groups of more and less effective programmes; and these results were fully replicated when 150 programmes were randomly split into two halves and analyses rerun on both sub-samples (see Appendix 2 in supplemental materials on publisher’s website). For 9-month outcomes, levels of significance for the full sample were $p < .001$, and improvement in correct classification beyond chance (i.e., 50%) was 32%. For 36-month outcomes levels of significance for the full sample were $p < .01$, and improvement in correct classification beyond chance was 35%.

**Do specific implementation ratings predict variation in programme effectiveness?**

To address this question, the 18 ratings were used in a stepwise regression with forward entry for each outcome, whereby the predictor with the strongest significant association was entered first, followed by additional predictors if they yielded a significant increase in predictive power. Given the substantial positive inter-correlation of the 18 ratings (mean .39, range .17 to .77), any significant effect may reflect the overall impact of ratings. Therefore, whenever a particular rating proved significant, the regression was repeated, adding the average of all 18 ratings (i.e., ratings’ composite) as an additional predictor. If this composite did not significantly affect the result, then the effect for an individual rating could be more confidently embraced.

One individual rating significantly predicted one of the two 9-month parenting outcomes and this held when the composite rating was included as a predictor; the more a SSLP promoted empowerment, the more it enhanced maternal acceptance ($\beta = .28$, $p < .01$).

With outcomes for 36-month-olds, programmes rated higher on identification of users exerted greater impact on children’s nonverbal ability ($\beta = .27$, $p < .01$), a result not changed by including the ratings’ composite as a predictor. Programmes rated higher on ethos and lower on service flexibility exerted more positive impacts on maternal acceptance, but the latter result appeared to be an artefact of statistical suppression between these two highly correlated variables ($r = .54$), as the latter proved unrelated to the outcome when considered on its own. When the composite rating was included as a predictor, the effect of ethos became insignificant, as did the composite itself, as a result of suppression between highly correlated predictors. Given the high correlation between ethos and the composite rating ($r = .80$), the significant effect of ethos ($\beta = .18$, $p < .05$) can be regarded as a proxy for a generally effective programme rather than an effect specifically attributable to ethos. Programmes exerted greater beneficial impact on maternal acceptance when programmes were rated higher on empowerment, which remained when the ratings’ composite was included as a predictor. Finally, the outcome home learning environment was significantly predicted by the rating of empowerment ($\beta = .34$, $p < .01$).

**Does service provision predict programme effectiveness?**

To address this question, the analysis first tested the relationship between an outcome and inherited services, and variables for improved and new services

---

**Table 2 Prediction of programme effectiveness at 9-months and 36-months by proficiency ratings based on discriminant function analysis**

<table>
<thead>
<tr>
<th>Predicted effectiveness</th>
<th>Bottom half</th>
<th>Top half</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed proficiency, 9-month-olds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottom half</td>
<td>49 (66%)</td>
<td>25 (34%)</td>
<td>74</td>
</tr>
<tr>
<td>Top half</td>
<td>26 (34%)</td>
<td>50 (66%)</td>
<td>76</td>
</tr>
<tr>
<td>Total</td>
<td>77</td>
<td>73</td>
<td>150</td>
</tr>
<tr>
<td>Observed proficiency, 36-month-olds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottom half</td>
<td>50 (67%)</td>
<td>25 (33%)</td>
<td>75</td>
</tr>
<tr>
<td>Top half</td>
<td>24 (32%)</td>
<td>51 (68%)</td>
<td>75</td>
</tr>
<tr>
<td>Total</td>
<td>74</td>
<td>76</td>
<td>150</td>
</tr>
</tbody>
</table>

9-months: Observed value ($H_o = \text{sum of main diagonal values} = 49 + 50 = 99$). Expected value ($H_e = \sum_{\text{row total}}^2 / N = (74^2 / 150) + (76^2 / 150) = 75$). $Z = (H_o - H_e) / \sqrt{N/\hat{\sigma}}$, $H_o - H_e = 3.92$ ($p < .01$). Improvement of discrimination over chance, $I = (H_o - H_e) / (N - H_o) \times 100 = 32\%$.

36-months: Observed value ($H_o = \text{sum of main diagonal values} = 50 + 51 = 101$). Expected value ($H_e = \sum_{\text{row total}}^2 / N = (75^2 / 150) + (75^2 / 150) = 75$). $Z = (H_o - H_e) / \sqrt{N/\hat{\sigma}}$, $H_o - H_e = 4.25$ ($p < .01$). Improvement of discrimination over chance, $I = (H_o - H_e) / (N - H_o) \times 100 = 35\%$.
were evaluated subsequently. In the hierarchical regression variables were allowed, within each stage, to enter stepwise with forward entry; thus variables only enter if they significantly improve prediction. For 9-month-olds, programmes that inherited more parent-focused services reduced negative parenting more ($\beta = -0.23, p < .01$). For 36-month-old parenting, the more child-focused services were improved, the more maternal acceptance increased ($\beta = 0.25, p < .05$).

**Do staffing patterns predict programme effectiveness?**

Staffing variables were analysed in the same way as service variables. Significant effects emerged for only one outcome, 36-month maternal acceptance. The greater the proportion of health services staff, the more maternal acceptance increased ($\beta = 0.26, p < .01$).

**Discussion**

An important step, for both theory and practice, is to move beyond evaluation of the overall impact of an intervention and identify factors and processes that might account for variation in impact, as is increasingly recognised for parent training programmes (Reyno & McGrath, 2006). The goal of the present study was to discover why some programmes may have been more effective than others in promoting child and family well-being, especially given the very modest overall effects of the SSLP initiative (Belsky et al., 2006). Results reveal a limited degree of linkage between variation in programme implementation and impact on children/families, offering some guidance for practice. The investigatory strategy adopted is novel and offers possibilities for application to other interventions where fidelity criteria are moot. This seems especially relevant given that it was based on the ‘theory of change’ approach to intervention evaluation (Weiss, 1995), which stresses the importance of developing interventions and assessing their implementation and impact in terms of their underlying theories or principles.

The relatively high number of associations tested raises questions about the confidence to be placed in the few significant results to emerge. Examining the pattern of results partly addresses this reservation, as the significant relationships between measures of implementation and of programme-specific impact on outcomes are all positive (i.e., higher implementation proficiency linked with more beneficial programme impact). If significant results reflected chance, some negative findings would be expected. The probability of the pattern of detected results occurring by chance for 8 significant findings is $0.5^8$, or $p < .004$ (and this probability decreases further if the split-half replications of discriminant analyses are considered). This suggests that results reflect more than chance and so the findings of this study of within-group differences are consistent with the between-group study results that SSLPs exert limited small effects in the desired direction (Belsky et al., 2006), especially when they are implemented in ways more consistent with the overall programme philosophy (as operationalised in the 18 proficiency ratings).

Though detected relationships are not strong, it is encouraging that programmes scoring higher on implementation proficiency also score higher on measures of programme impact on child and parenting outcomes. Collectively, 18 implementation-proficiency ratings differentiate more effective from less effective SSLPs for 9- and for 36-month-old child and parenting outcomes. In that proficiency ratings reflect adherence to the underlying principles in SSLP guidance, this result implies that programmes judged more proficient in putting guiding principles into practice were more likely to positively impact children and parenting. Perhaps it is not surprising that more proficient programmes are more effective. But these results are important for two reasons. First, the systematic linkage between proficiency measurements, especially all 18 in the replicated discriminant function analysis, serves to validate the otherwise only face-valid measures. Secondly, the proficiency findings illuminate the conditions under which this area-based programme can benefit children and families, whether or not it always or even routinely does so. Programmes tended to score similarly across all 18 ratings incorporating three broad aspects – what was implemented, the processes underpinning service delivery and holistic aspects of programme functioning. Conceivably, this is an artefact or ‘halo’ effect whereby a good rating on one or more dimensions biases the rater toward higher ratings on others. This is unlikely because every point on the 18 rating scales clearly defined what qualified for the relevant score – and because such high levels of inter-rater agreement were achieved. Proficiency in one domain goes with proficiency in other domains, and proficient implementation in general is more likely to produce better outcomes for children and families. For families, then, it matters not only what services are implemented but also that they are proficiently delivered. This requires a clear vision, cogent means of communicating that vision and a welcoming ethos.

Some effects attributable to specific programme features emerged. More empowerment in programmes was related to greater positive impact on maternal acceptance for mothers of 9-month-olds; better identification of users was related to more positive programme impact on nonverbal ability of 36-month-olds; stronger ethos and better proficiency overall were positively related to greater beneficial impact on maternal acceptance of 36-month-olds;
and more empowerment was related to greater success in fostering stimulating home learning environments for 36-month-olds. However, because the 18 ratings are inter-correlated, effects seemingly attributable to individual aspects of programme proficiency should be considered in relation to the overall effects of the composite measure of implementation proficiency, with higher overall proficiency predicting greater programme impact. Some may not be surprised by these findings. Yet a literature search was unable to find previous empirical evidence of the effects of such programme implementation characteristics upon children and families. Possibly this lack of evidence reflects the difficulties inherent in measuring such characteristics and this research offers innovative methods of measurement that could be adopted by others.

The finding that empowerment is related to two of the eight measures of programme impact upon child and parenting outcomes, in particular two of the five parenting measures (9-month-old maternal acceptance, 36-month-old home learning environment), suggests that strengthening programme activities relevant to empowerment should improve their effectiveness in influencing parenting. Note that empowerment in this study is not a self-report measure. Empowerment refers here to a rating of the extent to which procedures are in place that actually increase parent and staff participation and collaboration in decision-making and programme activities. The programme characteristics that go with empowerment include community groups and parents being involved in the planning and delivery of services; parent representation; staff training opportunities; clear exit strategies for users; services to include self-help groups; evidence that staff and users constitute a learning community; and evidence of mutual respect for all parties.

The enhanced effectiveness for parenting may promote child well-being, given evidence from several countries that maternal acceptance is related to more competent child functioning (e.g., Bradley, 2002) and research in the UK showing that the home learning environment has comparable beneficial effects (Melhuish et al., 2001, 2005). Additionally, in that intergenerational continuity for behaviour problems is at least partly mediated by intergenerational continuity in parenting (Smith & Farrington, 2004), improving parenting may be one way to disrupt cross-generational cycles of disadvantage, a central aim of SSLPs. Possibly links between socioeconomic disadvantage and later maladaptive behaviour may be reduced given evidence that it is adverse family functioning that mediates much of the long-term link between socioeconomic status in childhood and susceptibility to crime in young adulthood (Fergusson, Swain-Campbell & Horwood, 2004). With regard to the impact of programme services and staffing on effectiveness, significant results emerged only for 36-month-old parenting outcomes.

The fact that programmes with more inherited parent-focused services proved more successful in reducing negative parenting makes it clear that some benefits that might appear attributable to SSLPs may actually reflect service history in communities. The fact, though, that more improved child-focused services predicted greater positive programme impact on maternal acceptance makes it clear that a new area-based initiative can promote positive change when well implemented. Finally, the fact that this same programme impact (i.e., on maternal acceptance) was related to having a higher proportion of health-related staff is consistent with findings linking health-agency leadership to programme effectiveness (Belsky et al., 2006).

The importance of integrating health services in these multi-agency programmes is also suggested by the finding that those more proficient in identification of users had greater positive impacts on nonverbal ability in 36-month-olds. Health-led programmes were better positioned to identify families with young children, because of easy access to birth records (from health services), perhaps enabling early use of services that lead directly to stimulating experiences for children or, indirectly, via improvements in parenting. Also as health practitioners already had established systems for visiting families with young children prior to Sure Start, they could probably 'hit the ground running' when the intervention provided resources to improve services. The NESS Cost Effectiveness evaluation (Meadows, 2006) indicates that programmes led by health agencies became operational more quickly than those led by other agencies. Services that were implemented earlier by health staff (health visitors) with expertise in working with children 0–3 and their families were able to offer targeted treatment from birth and were therefore more likely to have a positive impact on parents, and ultimately their children.

Conclusions

As no intervention is likely to be universally effective, and evidence of features of comprehensive interventions that promote child and family well-being is sparse, it is critical to discover what elements of programmes account for variation in effectiveness. Such knowledge provides insight in how to make less effective programmes more effective. In order to inform practice, this study examined how implementation characteristics of a nationwide area-based intervention may influence variation in the impact on young children and their families living in disadvantaged communities. Multilevel modelling was used to estimate programme-specific effects. Then variation in implementation characteristics reflecting the theory of change underlying SSLPs, services and staffing was associated with variation in

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.
programmes’ collective impact on young children and their families. Linking these illuminated the conditions under which SSLPs had greater and lesser beneficial impacts on children/families. The resulting insights hold promise for practice and contribute to applied science by illustrating how diverse sources of often qualitative data can be quantified and subject to rigorous quantitative analysis.

It must be emphasised that other research indicates that it took around three years for these complex initiatives to approach full capacity in service delivery (Meadows, 2006), a period longer than originally anticipated by programme developers. While this study reveals some limited links between implementation and impact on children/families, understanding more clearly how different aspects of programmes are related to their impact may require closer scrutiny using case study methods. Results from this inquiry underscore that some of these community programmes were better functioning than others, at least in terms of realising the underlying their philosophy, and that this seems to matter with respect to their impact on children/families. This is an important message for the design of children and family services such as those to be delivered by Children’s Centres in the UK and for other similar community-wide interventions such as Stronger Families and Communities in Australia (FaCS, 2004). Also, for the evaluation of such interventions, theoretically derived ratings of proficiency may be a fruitful alternative to established measures of fidelity or quality.

Acknowledgements
Supported by a grant from the UK Department for Education and Skills.

Correspondence to
Edward Melhuish, Institute for the Study of Children, Families & Social Issues, Birkbeck, University of London, 7 Bedford Square, London WC1B 3RA, UK; Email: e.melhuish@bbk.ac.uk

Supplementary Material
The following supplementary material is available for this article:
Appendix S1. 18 ratings of implementation proficiency, as applied to Sure Start Local Programmes (SSLPs), with criteria for each point of scale.
This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1469-7610.2007.01705.x
Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors.

Any queries (other than missing material) should be directed to the corresponding author for the article.

References

© 2007 The Authors
Journal compilation © 2007 Association for Child and Adolescent Mental Health.


