

# Blueprints for Europe: Promoting Evidence-Based Programmes in Children's Services\*

## Blueprints para Europa: Promoviendo Programas Basados en la Evidencia en los Servicios de Atención a la Infancia

Nick Axford<sup>1</sup>, Delbert S. Elliott<sup>2</sup>, and Michael Little<sup>1</sup>

<sup>1</sup> The Social Research Unit, Dartington, UK

<sup>2</sup> University of Colorado, and Founding Director of Blueprints, USA

**Abstract.** This article describes the Blueprints database of evidence-based programmes (EBPs) and its potential application in children's services in European countries. It outlines relevant aspects of the European context, including a tendency to be skeptical about programmes imported from the US, and the need for a pan-European source of information about EBPs across multiple outcome areas. It then describes the standards of evidence used by Blueprints, which cover intervention specificity, evaluation quality, intervention impact, and dissemination readiness. The criteria for determining that a programme is 'Model' and 'Promising' are outlined. The article then summarizes the process by which the standards were developed and some of the issues that were harder to resolve. It also sketches the process by which a programme reaches the Blueprints database, and provides three examples of programmes approved by Blueprints and implemented in Europe: a home-visiting programme for mothers of infants; a parent skills training programme; and a therapeutic intervention for families of chronic offenders. A brief indication is also given of how the wider pool of programmes reviewed fare against the standards of evidence. Finally, the article summarizes future directions for the work, with a particular emphasis on how Blueprints might become widely used in Europe.

**Keywords:** child well-being, Europe, evidence-based program, program evaluation.

**Resumen.** En el artículo se describe la base de datos de Blueprints de programas basados en la evidencia (PBE's) y su aplicación potencial en servicios de atención a la infancia en los países europeos. Se abordan los aspectos relevantes del contexto europeo, y se señala una cierta tendencia al escepticismo ante programas importados de los EEUU, así como la necesidad de una fuente de información paneuropea sobre PBE's en todas las áreas de resultados. A continuación, se describen los criterios de evidencia empleados en los Blueprints, tales como la especificidad de la intervención, la calidad de la evaluación, el impacto de la intervención y la disponibilidad del sistema. Se examinan los criterios que se siguen para que un programa sea considerado como 'Modelo' y 'Prometedor'. A continuación, se resume el proceso mediante el que se desarrollaron los criterios y algunos de los aspectos de más difícil resolución. También se describen las líneas generales del proceso de incorporación de un programa a la base de datos de Blueprints, y proporciona tres ejemplos de programas aprobados por los Blueprints e implementados en Europa: un programa de visitas a domicilio a las madres de menores, un programa de entrenamiento de habilidades parentales, y una intervención terapéutica en familias de delincuentes reincidentes. También se indica brevemente que la mayor parte de programas revisados no cumple con los criterios de evidencia. Finalmente, el artículo resume las orientaciones futuras de trabajo en este campo, con especial énfasis en las distintas estrategias para extender el uso de Blueprints por toda Europa.

**Palabras clave:** bienestar infantil, Europa, evaluación del programa, programa basado en la evidencia.

The last decade has seen growing interest in evidence-based programmes (EBPs) in developed countries. A 'programme' is a discrete, organized package of practices, spelled out in guidance – sometimes called a

manual or protocol – that explains what should be delivered to whom, when, where and how. A programme is 'evidence-based' when it has been evaluated robustly, typically by randomized controlled trials (RCTs) or quasi-experimental designs (QEDs), and found unequivocally to have a positive effect on one or more relevant child outcomes (Social Research Unit, 2012a).

There are now many such programmes, covering all areas of children's lives, including education, behaviour, health, relationships and emotional well-being.

Correspondence: Nick Axford, Lower Hood Barn, Dartington, Totnes, TQ9 6AB Devon, UK. E-mail: [naxford@dartington.org.uk](mailto:naxford@dartington.org.uk)

\*Versión en castellano disponible en [spanish version available at]: [www.psyshosocial-intervention.org](http://www.psyshosocial-intervention.org)

However, their market penetration is very poor. They are rarely adopted, and, when they are, implementation is often poor and programmes fizzle out when initial funding ends (Bumbarger & Perkins, 2008). This holds in the US, where most EBPs originate, and even more so in Europe, where countries are increasingly importing these programmes.

There are various possible reasons for this state of affairs (Little, 2010). First, there is a lack of knowledge among policy makers and senior practitioners regarding the existence and nature of EBPs. Some arguably place little value on evidence. Second, there is confusion about what constitutes 'evidence-based' and what is an appropriate standard. Third, there is anxiety about whether these programmes can be implemented in real-world settings, and what human and financial resources this requires.

Such obstacles to implementing EBPs are magnified in Europe, where there is some resistance both to the concept of a programme and the fact that most of the best-known EBPs originate in the US (the two ideas tend to get conflated in debate). States with social-democratic or Catholic welfare regimes tend to be skeptical about programmes from the US (Grietens, 2010). This is partly because some programmes have had mixed or disappointing outcomes when implemented in Europe. For instance, *Multisystemic Therapy* performed only a little better than services as usual in the UK (Butler, Baruch, Hickey, & Fonagy, 2011), whereas in Sweden regular services did equally well (Sundell et al., 2008).

The hesitancy also reflects cultural differences in service provision between Europe and the US. The US has a minimal welfare state. North European countries, by contrast, have more redistributive policies and provide universal welfare (Rowlands, 2010). They also invest in professionalizing children's services staff, notably through social pedagogues who are trained in child development and work therapeutically with children and families in many settings (Petrie, Boddy, Cameron, Wigfall, & Simon, 2006). There is a reluctance to adopt practices from a country (the US) that routinely performs poorly in league tables of child well-being in developed countries, particularly since many European countries, especially those in Scandinavia, perform relatively well (UNICEF, 2007).

Compounding this skepticism, there is no pan-European source of information for a European audience on EBPs that cover a range of outcome areas. The best-known clearinghouses of EBPs are published in English and aimed primarily at American providers. Very few European programmes feature on them. Similar ventures in Europe tend to be country-specific,<sup>1</sup> or system-specific (e.g. education),<sup>2</sup> or focused on a single subject (e.g. drug prevention and treatment),<sup>3</sup> or concerned with a class of programmes (e.g. parenting interventions).<sup>4</sup> There is a recognized need for outlets in diverse languages and aimed at diverse cultures

(Soydan, Mullen, Alexandra, Rehnman, & Li, 2010). European children's services providers also need guidance on how to select and adapt programmes for a different context. This includes information on which children and families programmes succeed with.

This article discusses Blueprints, a resource that will provide policy makers and practitioners in Europe with high-quality information about programmes that meet a high standard of evidence and are ready for implementation in service systems. Blueprints is designed to lead to the greater awareness and use of EBPs and improved well-being for the children and families who receive them. We believe it fills an important niche. Before going any further it is helpful to provide some historical background.

Blueprints started in the US with a focus on violence prevention following the Columbine school massacre in 1993, in which two High School students killed 12 fellow students and a teacher. This was the catalyst for the Center for the Study and Prevention of Violence, part of the University of Colorado in nearby Boulder, to start compiling a list of evidence-based programmes specifically aimed at preventing violence (Elliott, 2010). In 2010, the Annie E. Casey Foundation funded the Social Research Unit (SRU) at Dartington, UK, and the Social Development Research Group (SDRG) at the University of Washington, US, to develop a method to help system leaders and communities to work better together to implement EBPs at scale. The idea was to bring together the centres' respective *Common Language* and *Communities that Care* methods (Axford & Morpeth, 2012; Hawkins & Catalano, 2002). The project necessitated the development of a menu of programmes that cover all key developmental outcomes, and so work began on a database of EBPs.

The researchers behind Blueprints for Violence Prevention were involved in developing the standards of evidence to underpin the *Evidence2Success* database (see below). They also recognized the value in broadening the scope of Blueprints, both in terms of outcomes and geography, hence the partnership with Casey and the Social Research Unit. With funding from Casey, Blueprints for Violence Prevention:

- Was re-named 'Blueprints for Problem Behavior and Healthy Youth Development', becoming the database for *Evidence2Success* (there is no separate *Evidence2Success* database).
- Essentially adopted the standard of evidence developed for *Evidence2Success* (which involved making slight changes to both the existing Blueprints standards and the *Evidence2Success* standards in order to align them).
- Broadened its remit to include programmes seeking to improve children's health, education, relationships, emotional well-being, and behaviour.
- Opened a European office in London, staffed by the Social Research Unit.

- Appointed its first European representative on the Board, and
- Re-designed its US website and started work on a European website (ensuring consistency in key content).

The remainder of this article describes the standards of evidence used to select programmes for the new Blueprints, how they were developed, the process by which a programme reaches the website, the kind of programmes that have been approved, how programmes fare against the standards, and future directions for the work with particular reference to Europe.

## The standards of evidence

The standards of evidence that underpin Blueprints today cover four dimensions:

1. *Evaluation quality* – whether the investigation into the efficacy and effectiveness of the programme produces valid and reliable findings.
2. *Intervention impact* – how much positive change in key developmental outcomes can be attributed to the programme.
3. *Intervention specificity* – whether the programme is focused, practical and logical.
4. *Dissemination readiness* – whether the programme is accompanied by the necessary support and information to enable its successful implementation in communities and public service systems.

Within each dimension, the Standards contain ‘Promising’ and ‘Model’ criteria. The Promising criteria set a basic minimum standard, while the Model criteria strengthen confidence in a programme’s scientific rigour, impact on outcomes, intervention specificity or readiness to be taken to scale. The four dimensions are now elaborated.

### Evaluation quality

In order for a programme to appear on Blueprints, it must have been evaluated by at least one good randomized controlled trial (RCT) or two good quasi-experimental design (QED) studies. ‘Good’ refers to aspects of methodological quality, specifically that:

- The method of assignment to intervention and control is at the appropriate level (eg. individual, school).
- The measurement instruments are suitable for the intervention population of focus and desired outcomes.
- Analysis is based on ‘intent-to-treat’.
- The statistical procedures and tests are appropriate.
- Intervention and control groups are equivalent at

baseline on key outcomes or appropriate controls for differences are included in the analysis.

Such studies must also meet the following criteria. It must be clear with whom the programme was tested and what was actually received by the intervention and comparison groups. The measures used must be valid and reliable, capture a relevant outcome, and not be tied to the programme under scrutiny. In order to minimize bias, someone *without* a vested interest in the programme must have provided the observations, ratings or assessments.

The extent to which participants dropped out during the study is also considered. Some drop-out is common, but it is problematic if many youth drop out, if some categories of youth drop out a lot more than others (eg. boys more than girls), or if the drop-out rate and type of person dropping out differ significantly between the programme and control groups.

There are several Model evaluation quality criteria. One is simply if there are at least two RCTs or one RCT and one QED that meet the quality criteria. Generally there can be more confidence in findings if a programme has been evaluated well more than once. Many evaluations only look at impact at the end of the programme, which is a problem as impact often fades with time. For this reason, evidence of a longer-term effect – at least 12 months after the programme ends – is also credited. These two criteria must be met for certification as a Model Programme.

Signs that evaluators have sought a finer-grained understanding of programme impact are also recognized. For instance, they might have tested for the relationship between implementation fidelity and outcome, or between the amount of programme received and outcome. If a programme has been delivered well, or if some youth or families have had more of it, a stronger effect would be expected. Some studies examine whether the programme works better for some sub-groups of than others, focusing on gender, ethnicity and socio-economic status. Studies may also test whether the logic that underpins the programme actually holds up; do effects take place for the reasons that were expected? Both of these are also Model criteria.

### Intervention impact

Programmes that appear on Blueprints must have a positive effect on a relevant outcome, the size of which is known, and have no known harmful effects. Only evaluation studies that meet all Promising evaluation quality criteria may be considered when making this judgment.

A majority of studies that meet this threshold must show that the programme has a positive effect on a relevant outcome in order for it to be judged to have an overall positive effect. A ‘positive effect’ means that

programme group youth or families did better relative to youth or families in the comparison group. It is important that this effect is not likely to be the result of chance, so it has to be 'statistically significant'. The size of this positive effect should also be provided.

There should be *no* evidence of the programme having a harmful effect on youth or families. This includes all outcomes and all sub-groups. For example, a programme would not be approved if it improves adolescents' relationships with peers but at the expense of their use of illicit substances increasing. Equally, if the programme decreases substance use for boys but not girls, it would be approved for boys only.

There are two Model criteria for Intervention impact. One is the existence of several studies meeting the Promising evaluation quality criteria, a majority of which show a statistically significant positive impact. The other is evidence that children who received a larger amount of the programme did better than those who received a smaller amount: in other words, there is a positive dose-effect relationship.

### ***Intervention specificity***

Programmes need to be clear about what outcomes they target and which group of children will benefit. There should be a clear description of what the programme comprises, and an explanation of why and how the programme should work – in other words, how the programme will address the risk and protective factors as a means of achieving the outcomes.

There is one Model criterion for Intervention specificity, which is the existence of compelling research evidence to support the programme logic. This must explain why and how the programme is likely to benefit the children and youth it is aimed at. For example, if a parenting programme encourages parents to practise certain skills to deal with their children's poor behaviour, have other studies shown that doing this works?

### ***Dissemination readiness***

Programmes that are accepted for the database also need to demonstrate that they can be implemented at scale in communities and service systems. At the simplest level, the programme that was evaluated should still be available. Next, it should be clear how to get the programme to the right children, youth and families. A manual and training and implementation materials are also needed, because these will help ensure that the programme is implemented consistently (or with fidelity). The financial and human resources needed for implementation should be stated.

There are several Model dissemination readiness criteria, starting with the availability of technical sup-

port with implementation and a checklist to help monitor fidelity. Recognition is given to programmes that are currently being disseminated widely, or that have been tested and found effective when delivered by regular practitioners in normal settings. Many programmes are tested initially under special conditions – for example, they are delivered in university clinics by research staff. Policy makers can have more confidence in programmes that have been tested when delivered by the kinds of people who normally provide similar services in their daily work.

### **Developing the standards**

In their work on *Evidence2Success*, the SRU, SDRG and Annie E. Casey Foundation recognized that there are already over 25 databases of evidence-based programmes and, accordingly, several sets of standards.<sup>5</sup> In an attempt to build some consensus, the decision was taken to develop the *Evidence2Success* standards of evidence – since adopted by Blueprints with some amendments, as described above – by involving experts who had previously developed other sets of standards of evidence, all of which but the last listed below have been used to inform databases of EBPs:

- Best Evidence Encyclopedia<sup>6</sup> (Robert E. Slavin, Johns Hopkins University, US).
- Blueprints for Violence Prevention<sup>7</sup> (Delbert S. Elliott, University of Colorado, US).
- LINKS (Lifecourse Interventions Nurturing Kids Successfully)<sup>8</sup> (Kristen Moore, Child Trends, US).
- Communities that Care<sup>9</sup> (J. David Hawkins and Richard F. Catalano, SDRG, University of Washington, US).
- Greater London Authority Project Oracle<sup>10</sup> (Michael Little, SRU at Dartington, UK).

These experts met regularly over a six-month period and tested prototype standards empirically to see how easy they were to apply and which programmes would meet them. Consideration was also given to other sets of standards, such as those developed by the Society for Prevention Research (Flay et al., 2005) and the CONSORT guidelines on reporting RCTs.<sup>11</sup> During this period several issues arose that required discussion. Most concerned evaluation quality and intervention impact. Some of the more important ones are now outlined briefly.

There was discussion about whether the amount of attrition from a study was important and whether a level at which it becomes problematic should be specified. However, it was argued that this would penalise follow-up studies, which the standards encourage, since these tend to have higher attrition. The expert group therefore decided to focus mainly on *differential* rather than overall attrition, although the Blueprints

Board still considers attrition rates and expects some controls or adjustment (e.g. propensity scoring) when this rate is high.

Another issue concerned the value of independent replication, in other words a study not involving the programme developer that nevertheless shows an impact. The rationale was that developer involvement seems to introduce bias (Eisner, 2009). The expert group agreed that while independent replication is desirable, and should be encouraged, to insist on it now would result in a very short list and remove stronger programmes, such as Nurse Family Partnership.

There was considerable discussion about the acceptability of evidence of impact derived solely from self-report measures. In criminal justice, such measures are generally considered acceptable. In education, they are unacceptable if used alone; observations, teacher ratings, and academic test scores are preferred. The expert group decided that the focus should be on whether measures are appropriate: no form of measurement is wrong *per se*. They insisted, however, that assessments cannot be limited to those made by persons who are not blind to the experimental condition or who are providing the intervention, since this can introduce bias.

There was also extensive debate over the requirement for evidence of sustained impact 12 months after the intervention ended. The worry was expressed that very few educational programmes would qualify because the last measure is usually taken at the end of the intervention. Evidence of a sustained impact was therefore made a 'best' rather than a 'good enough' criterion.

The main issue on intervention impact concerned the requirement for a statistically significant effect size. It was argued that in many studies with large clusters (such as schools) it is difficult to obtain large enough sample sizes to permit analysis at the cluster level. However, unbiased and meaningful estimates can be obtained using participant-level analyses when sample sizes at the participant level are large. The expert group therefore agreed to have as an alternative a sample size weighted mean effect size of 0.2, with a sample size of more than 500 individuals across all studies<sup>12</sup>.

Finally, a decision needed to be made about the criteria for determining a programme's overall status. Blueprints categorizes programmes as either 'Model' or 'Promising'. It was agreed that a programme must meet all 'good enough'<sup>13</sup> criteria across all four dimensions to be deemed 'Promising' and that a programme is designated 'Model' if it meets these criteria and:

- It has (a) two or more good enough randomised controlled trials or (b) at least one good enough randomised controlled trial and one good enough quasi-experimental design evaluation, *and*
- There is evidence of a sustained impact (at least 12 months after the end of the programme).

## How programmes get onto Blueprints

These standards are applied to programmes that seek to achieve outcomes in the areas of behaviour, emotional well-being, educational skills and attainment, health (particularly as it relates to behavioural issues, such as smoking, eating, drinking), and relationships (primarily with parents and peers). There are four steps in the process for Blueprints to approve a programme.

First, all relevant scientific literature on the programme is identified. (At present this is restricted to English-language publications.) The research team sifts through the primary journals in all areas of problem behaviour and child health and development on a regular basis to identify literature that might suggest new programmes for inclusion or add to or challenge programmes on the database. Information submitted by programme developers or purveyors is also considered.

Second, this literature is analyzed against the standards of evidence by a team of trained reviewers based at the University of Colorado in the US and the Social Research Unit in the UK. The result is a structured narrative description of each study and a quantitative summary of whether overall the programme meets each of the criteria contained within the standards. The reviews focus primarily on intervention specificity, evaluation quality and intervention impact. Each review must be approved for quality by a review coordinator.

Third, programmes that are deemed to have a good chance of meeting the standards of evidence are forwarded to the Blueprints Board for consideration. The Board comprises eight leading prevention scientists from the US and Europe and meets twice a year. The Board decides whether or not programmes meet the standards in terms of evaluation quality and intervention impact and can therefore potentially be recommended for dissemination.

Fourth, the review team checks the system readiness of programmes approved by the Blueprints Board. This is done by consulting programme websites and by asking developers or purveyors to complete a written questionnaire. The questionnaire covers subjects such as the availability of materials and training, fidelity monitoring procedures, and human resource requirements. If extra information is needed once this is submitted then follow-up questions are sent and, often, a telephone discussion is held. Two members of the review team – at least one with extensive experience of delivering and managing services – discuss the information received and determine if the programme is 'system ready'. At this stage a programme is formally approved and the developer is informed that it will appear on the Blueprints website.

The list of approved programmes is updated regularly. Regular literature searches are undertaken using a consistent process to identify new studies showing

positive or even negative findings for programmes already on the list. Similarly, studies on new programmes are reviewed if they seem likely to meet the standards of evidence. Particular efforts are being made to identify programmes originating in Europe since most programmes approved to date were developed in the US. Programme developers and evaluators may also submit their programme for consideration.

### Programmes approved for Blueprints

At the time of writing there are 11 'Model' programmes and 22 'Promising' programmes on Blueprints. This is from over 1000 programmes reviewed. However, several new programmes will shortly be added to the list, largely as the result of the decision described earlier in this article to extend the remit of Blueprints beyond *violence* prevention (the initial focus) to encompass other areas of child and youth development, such as education and health.

Some examples of programmes appearing on the Blueprints website follow. They represent different types and levels of intervention and have all been implemented in Europe. The descriptions outline briefly how each of the programmes meets the standards and where they are delivered.

Nurse Family Partnership is a home-visiting programme that involves nurses making home-visits to young, often teenage, vulnerable first-time parents, starting in early pregnancy and lasting until children are 24 months old. The programme aims to promote prenatal health, improve child well-being and development through better parenting, and encourage parental self-sufficiency through education, employment, or planning future pregnancies. Specially trained nurses pay weekly or fortnightly structured home-visits to families. Home-visits allow nurses to prepare young people for parenthood and guide them to adopt healthier lifestyles, take good care of their babies, and plan for their future. Key to the programme is the strong therapeutic relationship built between nurse and family.

Rigorous scientific evaluations show that NFP leads to a range of improvements in child health and development, such as better child behaviour and academic achievement, more positive parenting practices, reductions in child maltreatment, and increased parental independence –including reduced welfare use (eg. Olds, Henderson, Chamberlin, & Tatelbaum 1986; Olds & Kitzman, 1990). These impacts are sustained long after the programme finishes, for example children in FNP are less likely to be involved in the juvenile justice system in adolescence (eg. Eckenrode et al., 2010).

Nurse Family Partnership is accompanied by an extensive package of support, including manuals, training and technical assistance. It has been implemented in the UK and the Netherlands. Every dollar

invested in the US version of the programme for low-income families yields a return of \$3.23 (Aos et al., 2011).<sup>14</sup>

The second programme described here, Incredible Years BASIC, is designed for parents of children aged 2-10 years with conduct problems. It seeks to improve family interaction and prevent early and persistent anti-social behaviour in these children.

The programme comprises a 12-week course of two-hour sessions delivered to a group of about 12 parents by two specially trained leaders. Parents are taught strategies to help them manage their child's problem behaviours, such as aggression, tantrums, and acting out. They also learn how to promote their child's social skills through emotion regulation. Sessions involve group discussion, videotape modelling, and the rehearsal of parenting techniques.

Incredible Years BASIC has been evaluated by RCT in several countries, including the US, UK, and Norway. These evaluations show consistently that the programme increases the use of positive parenting strategies, reduces the use of harsh and inconsistent discipline, and reduces deviant behaviour in children (eg. Webster-Stratton, Kolpacoff, & Hollinsworth, 1988; Hutchings et al., 2007; Larsson et al., 2009; Scott et al., 2010; Little et al., 2012; McGilloway et al., 2012).

Incredible Years BASIC has extensive group leader manuals, DVDs, books, CDs, handouts, and recommended activities and reading between sessions. Group leaders receive initial three-day training and ongoing technical support and supervision to assist successful implementation. The programme has been provided in mental health agencies, public health centres and schools in the US, UK, Ireland, Norway, Germany, Denmark, Netherlands, Norway, Portugal and Sweden. For every \$1 invested the programme produces a return of \$1.20 (Aos et al., 2011).<sup>15</sup>

The third programme, Multisystemic Therapy (MST), is an intensive family-based intervention for adolescents who are chronic offenders; typically they have committed serious crimes and have substance abuse problems. MST aims to reduce anti-social behaviour and criminal activity, as well as improve parenting skills, family relations, school grades and involvement with positive peers and activities. A therapist works with the adolescent in their daily surroundings – with their family, friends, at school and in their community. Together with the family, the therapist designs a treatment plan to tackle identified risks and encourage protective influences in the adolescent's environment. Various strategies are employed, such as CBT or coaching. The therapist becomes a single point of contact for the family, available 24/7. A typical MST intervention lasts 3-5 months and involves 3-6 sessions weekly, each up to two hours long.

MST has been proven to work in multiple high-quality RCTs. It reduces criminal recidivism rates and anti-social behaviour, including conduct problems and

aggression, and also improves emotion management and family cohesion (eg. Henggeler, Melton, & Smith, 1992; Timmons-Mitchell, Bender, Kishna, & Mitchell, 2006). Some effects are long-lasting, with improvements still visible several years after treatment (eg. Henggeler, Clingempeel, Brondino, & Pickrel, 2002).

A US-based organisation 'MST Services' provides training, technical support, monitoring, materials including treatment manuals, and licensing. It is delivered by experienced therapists, each of whom receive five days training and ongoing support. MST has been delivered extensively in the US, and in several European countries, including Norway, Spain, Sweden, Denmark, Netherlands, Iceland and the UK. A UK cost-benefit analysis reveals that every pound spent on MST produces a return of £1.77 (Social Research Unit, 2012b)<sup>16</sup>.

### How programmes fare against the standards of evidence

It is instructive to reflect briefly on how programmes perform against the standards of evidence. Here the focus is on the 100 programmes reviewed against the standards in 2011 as part of the Annie E. Casey Foundation *Evidence2Success* project. These represent a spread of programmes in terms of child developmental stage and outcomes targeted, and were deliberately selected because the expert group deemed them to be the best available.

*Intervention specificity* was generally good, with most programmes reviewed meeting each of the 'good enough' criteria on this dimension. *Evaluation quality* was much more variable. The 'good enough' evaluation quality criteria that tended to be better addressed in programme evaluations include: the appropriateness of measures (reflecting outcomes, not being tied to intervention, not being rated solely by the implementer); having a clear statement of demographics; and assigning cases to programme and comparison groups at the appropriate level (although they are not always analyzed at the correct level). There was less clarity about: what the control group received; how the intervention that was actually delivered compares with intervention as it was designed to be delivered; if or how clustering is controlled for in analyses, for example when the unit of allocation is schools; whether there is equivalence between the programme and control groups at baseline on outcome measures; whether analysis was intent-to-treat or not; and whether there was differential attrition.

Regarding the 'Best'<sup>17</sup> evaluation quality criteria, 12-month follow-up was available in fewer than half of cases, as was sub-group analysis. It was rare for there to be any analysis of the relationship between fidelity and outcomes or of the role played by mediating fac-

tors. Dose-response analysis – in the proper sense of setting out deliberately to vary the dose and comparing, for example, a full-length version and a shorter version of the programme – was extremely rare.

On *intervention impact*, about half of the programmes indicated effect size. As regards *system readiness*, it was generally difficult to establish system readiness without contacting the programme developer; the information supplied on programme websites, for example, is inadequate for that purpose. Usually the programme that was evaluated is still available, although this can be difficult to detect as programmes 'morph' over time, notably to make improvements or to be more suitable for a different population or setting. Most programmes reviewed had a manual and training but information about financial and human resources was much less readily available. The extent of dissemination and 'real world' testing was often unclear, and although many programmes purport to have a fidelity protocol it was less clear whether this is suitable for use beyond research studies, in other words in orthodox service settings.

### The Blueprints websites

Until now, the work of Blueprints has been disseminated through a website managed by the Blueprints for Violence Prevention team at the University of Boulder Colorado. The new Blueprints will have two websites. The main site is aimed primarily at a US provider audience. It is designed and maintained to enable policy makers and providers to access readily the information they need on each approved programme. The approach taken is similar to that used in 'Consumer Report' (in the US) or 'Which?' (UK) magazines. Instead of searching for and comparing cameras or washing machines, website users can search for suitable programmes by outcome, target group, and risk and protective factors. It is not possible here to give an exhaustive list of the information supplied about each programme but the main fields include:

- Programme objectives.
- Programme recipients.
- Level of intervention (eg. universal prevention, selected prevention, treatment).
- Setting (eg. school, community, home).
- Targeted risk and protective factors.
- A brief description of the programme.
- A brief description of the outcomes achieved by the programme.
- A brief description of the methodology used in the relevant evaluation studies.
- Financial information (eg. unit cost, cost-benefit ratio, potential funding strategies).
- Training and technical assistance information.
- Contact information for both the programme designer and the purveyor.

Users can print a fact sheet containing this information and also compare different programmes against the same criteria.

A sister website aimed at European providers will have essentially the same functions and be consist in content, but it will contain less text, include a handful of different fields (eg. evidence of dissemination of the programme in question in Europe) and be translated into European languages (initially Spanish, French and German). Users wishing to obtain further information will be able to link to the US website.

Both websites are ongoing projects. It is planned in due course to add new fields, including a visual logic model, video content (eg. the programme developer summarizing the programme), the facility to explore the likely impact of implementing a *portfolio* of programmes on costs, benefits and outcomes, and subjective feedback from policy makers, practitioners, children and families who have experience with the programmes.

The task of producing the websites requires generating high-quality content for each approved programme. Some of this comes direct from the completed reviews, including outcomes targeted, target group, and logic model. However, additional data collection is required. First, a consistent and comparable indicator of the size of effect will be generated using the meta-analytic methods in operation at the Washington State Institute for Public Policy (Aos et al., 2011; Lee, Drake, Pennucci, Bjornstad, & Edovald, 2012). In the case of programmes and outcomes not yet examined by the Washington centre this requires coding studies, conducting meta-analyses and applying the effect size formula utilized by the Washington State Institute for Public Policy. Second, financial data is also created on the cost of implementation. This is broken down into costs for start-up, materials, delivery, training and technical assistance. Potential strategies and sources for funding the programme are also listed. This information is obtained via questionnaires and interviews with programme developers/purveyors. Third, the Washington centre uses the effect size and unit cost data to generate a cost-benefit ratio, which will be used by Blueprints.

### Future developments

The work described here will develop in at least six ways. First, it is expected that the Blueprints standards, which already set a high bar in the child welfare field, will become higher as understanding of the science of evidence-based programmes and their implementation improves and as the quality of studies improves in response to standards such as these. For example, the stipulation of sustained impact at 12-month follow-up might move from being a Model criterion to become a Promising criterion, and the requirement for an independent replication might be added.

Second, continuing efforts will be made to build understanding of how the Blueprints standards and database link to those used by other groups. For example, there will be ongoing discussions with groups such as the Society for Prevention Research, and in Europe work has started on building connections with research and intervention communities in participating countries – including those representing existing clearinghouses. Such collaboration will contribute to the wider use of Blueprints *and* other sources. For example, some clearinghouses, such as MOVISIE in the Netherlands, track innovation, whereas Blueprints focuses on proven programmes. Both are important, and the work of the former should contribute to the latter.

Third, new programmes will be approved. More programmes developed in Europe need to be rated against the standards of evidence. Several that stand a reasonable chance of meeting the standards are known of but they need to be reviewed in full and, if approved by the Blueprints Board, disseminated more widely (eg. Atria & Spiel, 2007; Faggiano et al., 2008; Salmivalli, Kärnä, & Poskiparta, 2011). In addition, and with a view to the longer-term, work is underway to show how the standards can be used to help practitioners to take programmes on the journey from innovation to model programme. This argues that innovations should be strengthened and tested with a level of rigor appropriate to their stage of gestation (Little, 2012a). For example, new programmes might warrant a pre-post or even a small comparison group study, with progress to larger RCTs conditional on positive results.

Fourth, Blueprints will need promoting in Europe. In the first instance, the standards of evidence should be shared with colleagues in relevant European organizations, including pan-European centres, clearinghouses and country-specific research institutes. This will promote discussion about the differences and similarities between different standards/databases (e.g. in terms of focus, function and audience) and ways of connecting different initiatives to achieve greater synergy. Eventually, specified European states and/or the European Union might be encouraged to adopt the standards, database and economic model. The strong alliances forged by Blueprints with key experts and decision makers in the US need to be replicated in Europe. This will require written materials as well as conferences and meetings to understand potential users' needs and concerns and encourage them to use the website. Training materials on the standards and using the websites might be developed for end-users.

Fifth, research is needed on the transportability of programmes from one context to another – in particular from the US to Europe, or from one European country to another. The success of imported programmes is mixed, as indicated already. What is effective in one context might not be effective in another, and what is



culturally appropriate in one context may not be in another. Various factors might account for this, including the extent to which programmes need adaptation and are adapted carefully (Kumpfer et al., 2012) and the policy and service delivery context of the new site (Sundell et al., 2008). But more research is needed, and European providers need information about whether imported programmes are likely to 'fit' or work in their country and how the chances of this happening can be improved. Further, the costs and benefits of programmes may differ across countries due to differences in welfare systems, hence the importance of the translation of the Washington economic model referred to above.

Lastly, Blueprints focuses on programmes but there is growing interest in the idea of kite-marking policy and in identifying effective *practices* (eg. Barth et al., 2012), given the difficulties and limitations often of implementing programmes in children's services systems (Little, 2010). For this reason, new standards of evidence will be developed and used to identify evidence-based practices, policies and processes (eg. assessment methods) that should be recommended for widespread dissemination.

## References

- Aos, S., Lee, S., Drake, E., Pennucci, A., Klima, T., Miller, ... Burley, M. (2011). *Return on investment: Evidence-based options to improve statewide outcomes. Document No. 11-07-1201*. Olympia: Washington State Institute for Public Policy.
- Atria, M., & Spiel, C. (2007). The Viennese Social Competence (ViSC) training for students: Program and evaluation. In J. E. Zins, M. J. Elias, & C. A. Maher (Eds.), *Bullying, victimization and peer harassment: A handbook of prevention and intervention*. (pp. 179-198). New York, NY: Haworth Press.
- Axford, N., & Morpeth, L. (2012). The Common Language prevention operating system: from strategy development to implementation of evidence-based practice. In B. Kelly & D. Perkins (Eds.), *Handbook of Implementation Science for Educational Psychology*. Cambridge: Cambridge University Press.
- Barth, R. P., Lee, B. R., Lindsey, M. A., Collins, K. S., Strieder, F., Chorpita, B. F., ... Sparks, J. A. (2012). Evidence-based practice at a crossroads: The emergence of common elements and factors. *Research on Social Work Practice*, 22, 108-119.
- Bumbarger, B. K., & Perkins, D. F. (2008). After randomised trials: issues related to the dissemination of evidence-based interventions. *Journal of Children's Services*, 3, 55-64.
- Butler, S., Baruch, G., Hickey, N., & Fonagy, P. (2011). A Randomized Controlled Trial of Multisystemic Therapy and a statutory therapeutic intervention for young offenders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 50, 1220-1235.
- Eckenrode, J., Campa, M., Luckey, D. W., Henderson, C. R., Cole, R., Kitzman, H., ... Olds, D. (2010). Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized trial. *Archives of Pediatric Adolescent Medical*, 164, 9-15.
- Eisner, M. (2009). No effects in independent prevention trials: can we reject the cynics' hypothesis? *Journal of Experimental Criminology*, 5, 163-183.
- Elliott, D. S. (2010). Lessons from Columbine: effective school-based violence prevention strategies and programmes. *Journal of Children's Services* 4, 53-62.
- Faggiano, F., Galanti, M. R., Bohrn, K., Burkhart, G., Vigna-Taglianti, F., Cuomo, L., ... Wiborg, G., EU-Dap Study Group (2008). The effectiveness of a school-based substance abuse prevention program: EU-Dap cluster randomised controlled trial. *Preventive Medicine*, 47, 537-543.
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., ... Peter, J. I. (2005). Standards of Evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151-175.
- Grietens, H. (2010). Discerning European perspectives on evidence-based interventions for vulnerable children and their families. *International Journal of Child and Family Welfare*, 13, 6-17.
- Hawkins, J. D., & Catalano, R. F. (2002). *Communities that Care - Tools for Community Leaders: A Guidebook for Getting Started*. South Deerfield, MA: Channing-Bete.
- Henggeler, S. W., Melton, G. B., & Smith, L. A. (1992). Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology*, 6, 953-961.
- Henggeler, S. W., Clingempeel, W. G., Brondino, M. J., & Pickrel, S. G. (2002). Four-year follow-up of multisystemic therapy with substance-abusing and substance-dependent juvenile offenders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 868-874.
- Hutchings, J., Bywater, T., Daley, D., Gardner, F., Jones, K., Eames, C., ... Whitaker, C. (2007). A pragmatic randomised control trial of a parenting intervention in sure start services for children at risk of developing conduct disorder. *British Medical Journal*, 334, 678-682.
- Larsson, B., Fossum, B., Clifford, G., Drugli, M., Handegard, B., & Mørch, W. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children: results of a randomized controlled trial. *European Child & Adolescent Psychiatry*, 18, 42-52.
- Lee, S., Drake, E., Pennucci, A., Bjornstad, G., & Edovald, T. (2012). Economic evaluation of early childhood education in a policy context. *Journal of Children's Services*, 7, 53-63.
- Little, M. (2010). *Proof Positive*. London: Demos.
- Little, M. (2012). *From Innovation to Proven Model*. Dartington: Social Research Unit.
- McGilloway, S., Mhaille, G., Bywater, T., Furlong, M., Leckey, Y., Kelly, P., ... Donnelly, M. (2012). A parenting intervention for childhood behavioural problems: A ran-

- domized controlled trial in disadvantaged community-based settings. *Journal of Consulting and Clinical Psychology*, 80, 116-127.
- Olds, D. L., Henderson, C. R., Chamberlin, R., & Tatelbaum, R. (1986). Preventing child abuse and neglect: A randomized trial of nurse home visitation. *Pediatrics*, 78, 65-78.
- Olds, D. L., & Kitzman, H. (1990). Can home visitation improve the health of women and children at environmental risk? *Pediatrics*, 86, 108-116.
- Petrie, P., Boddy, J., Cameron, C., Wigfall, V., & Simon, A. (2006). *Working with children in care: European perspectives*. Maidenhead: Open University Press.
- Rowlands, J. (2010). Services are not enough: child well-being in a very unequal society. *Journal of Children's Services*, 5, 80-88.
- Salmivalli, C., Kärnä, A., & Poskiparta, E. (2011). Counteracting bullying in Finland: The KiVa program and its effects on different forms of being bullied. *International Journal of Behavioral Development*, 35, 405-411. doi:10.1177/0165025411407457
- Social Research Unit (2012a). *An introduction to evidence-based programmes in children's services*. Dartington: Social Research Unit.
- Social Research Unit (2012b). *Investing in Children: Youth Justice 1.1*. Dartington: SRU.
- Scott, S., O'Connor, T. G., Futh, A., Matias, C., Price, J., & Doolan, M. (2010). Impact of a parenting program in a high-risk, multi-ethnic community: The PALS trial. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 51, 1331-1341.
- Soydan, H., Mullen, E. J., Alexandra, L., Rehnman, J., & Li, Y. P. (2010). Evidence-based clearinghouses in social work. *Research on Social Work Practice*, 20, 690-700.
- Sundell, K., Hansson, K., Lofholm, C. A., Olsson, T., Gustle, L. H., & Kadesjo, C. (2008). The transportability of Multisystemic Therapy to Sweden: Short-Term results from a randomized trial of conduct-disordered youths. *Journal of Family Psychology*, 22, 550-560.
- Timmons-Mitchell, J., Bender, M., Kishna, M. A., & Mitchell, C. (2006). An independent effectiveness trial of Multisystemic Therapy with juvenile justice youth. *Journal of Clinical Child and Adolescent Psychology*, 35, 227-236.
- Webster-Stratton, C., Kolpacoff, M., & Hollinsworth, T. (1988). Self-administered videotape therapy for families with conduct-problem children: Comparison with two cost-effective treatments and a control group. *Journal of Consulting Clinical Psychology*, 56, 558-566.

## Footnotes

<sup>1</sup> Examples include SFI Campbell (Denmark), MOVISIE (Netherlands), Ungsinn (Norway), Metodguiden and SBU (Sweden), and Prevención basada en la evidencia (Spain).

<sup>2</sup> For example, the Best Evidence Encyclopaedia (UK).

<sup>3</sup> For example, the best practice portal of the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), known as the European Exchange on Drug Demand Reduction Action. (EDDDRA).

<sup>4</sup> For example, the National Academy of Parenting Research (NAPR) Parenting Programmes Evaluation Tool (UK).

<sup>5</sup> The authors can supply a list of these if required.

<sup>6</sup> <http://www.bestevidence.org/>

<sup>7</sup> <http://www.colorado.edu/cspv/blueprints/>

<sup>8</sup> <http://www.childtrends.org/links/>

<sup>9</sup> The Prevention Strategies Guide that was informed by the Communities that Care standard is no longer available. CTC specifically direct sites to Blueprints.

<sup>10</sup> <http://www.london.gov.uk/priorities/crime-community-safety/time-action/project-oracle>

<sup>11</sup> <http://www.consort-statement.org/>

<sup>12</sup> This was agreed to as a temporary provision until it could be determined how it affected the ratings of educational programmes. Since the review process indicated that educational programmes were *not* being rejected on the basis of the statistical significance requirement, the Blueprints Board does not use the provision.

<sup>13</sup> 'Good enough' criteria in *Evidence2Success* are essentially 'Promising' criteria in Blueprints (with some modifications).

<sup>14</sup> Applicable when delivered in Washington state.

<sup>15</sup> Applicable when delivered in Washington state.

<sup>16</sup> Applicable when the programme is delivered in England and Wales.

<sup>17</sup> 'Best' criteria in *Evidence2Success* are essentially 'Model' criteria in Blueprints (with some modifications).

Manuscript received: 04/10/2011

Review received: 07/12/2011

Accepted: 09/12/2011