

This is a pre-copyedited, author-produced version of an article accepted for publication in The British Journal of Social Work following peer review. The version of record Hood, Rick (2019) What to measure in child protection? The British Journal of Social Work, 49(2), pp. 466-484. is available online at: <https://academic.oup.com/bjsw/article/49/2/466/5062282>.

# What to measure in child protection?

---

## **Abstract**

This paper explores key conceptual frameworks for measurement in child protection services, and considers their application in the English statutory system. After introducing some of the debates surrounding the use of measures in child protection, three different perspectives are considered: performance-based accountability, evidence-based approaches, and socio-technical systems design. The paper outlines the main principles of each perspective and their implications for measurement, drawing on examples from the relevant literature. It is argued that the merits and drawbacks of different measures are dependent on the conceptual frameworks in which they are used, and these in turn reflect the contested nature of child protection and its institutional context.

Keywords: child protection, performance management, measures, children's services, systems theory

## Introduction

Recent decades have seen increasing interest in the measurement of public services, driven by a combination of economic, administrative and political considerations. An underlying concern has been with productivity; the Office for National Statistics (ONS) articulates the premise – implicitly or explicitly accepted by all governing parties since 1979 – that ‘increasing productivity is generally considered to be the only sustainable way of improving living standards in the long term’ (ONS, 2016: 4). Since the public sector constitutes just over a fifth of the UK economy, measuring inputs and outputs in order to derive indicators of productivity has become a staple of governance. During the 1980s, the idea of measuring public services was refracted through the ideology of New Public Management (NPM), which sought to make the professional bureaucracies of the welfare state transparent and accountable to the citizens who use them (Levy, 2010). The econometric concept of productivity has therefore become associated with its technocratic correlates: efficiency, meaning value for taxpayer’s money, and effectiveness, meaning empirical evidence of success. The process of maximising efficiency and effectiveness in the public sector, drawing on managerial methods imported from the private sector, has become known as performance management (Van Dooren et al., 2015). In this guise, productivity measures have become vital to the regulatory role of many public services, which not only provide a service to the citizen-consumer but also tackle social problems such as crime, pollution, and abuse (Sparrow, 2000). Measures have therefore come to serve a dual purpose within contemporary ‘risk regulation regimes’ (Hood et al., 2001), firstly in helping to identify and mitigate societal risks and secondly in helping to defend against the institutional risk of criticism and scandal when things go wrong.

In the centralised political culture of England, child protection (CP) and welfare services for children and families (sometimes collectively referred to as child safeguarding or children’s social care) have closely mirrored these developments. The influx of performance management has led to a proliferation in audit and quality assurance mechanisms in statutory social services (Munro, 2004), underpinned by statistical returns generated by electronic workflow systems (Shaw et al., 2009). External inspection of these services was taken over in 2007 by the Office for Standards in Education, Children’s Services and Skills (Ofsted), which has taken a robust view of performance across the sector. These changes have been

controversial, as might be expected when command-and-control management encroaches on domains where professionals have previously been in charge of the work (Freidson, 2001). At the same time, attention has been drawn to trends in demand and provision over the same period, notably the acceleration in CP interventions (Bilson and Martin, 2016; Hood et al., 2016a) and the existence of welfare inequalities linked to variations in local deprivation and expenditure (Bywaters et al., 2015).

The troubled institutional context of CP means that the question of what to measure is not a neutral or technocratic one; on the contrary, it reflects fiercely contested and politicised debates about purpose, organisation and method. As Forrester (2017) points out, an important question is who decides what the outcome of a service should be, or indeed who decides which children should receive a service and what kind of service they should receive? Issues around power and agency may be overlooked within a seeming consensus that services should be more 'outcome-focused'. The contribution of this paper is to examine what different measures tell us about the underlying (or unwitting) intentions of those who advocate or use them. In what follows, the title question will be considered from three different perspectives: performance-based accountability, evidence-based services, and socio-technical systems design.

## **Performance-based accountability (PBA)**

Performance-based accountability (PBA), also known as outcomes-based or results-based accountability (Friedman, 2001; Schorr, 1995), provides a conceptual framework for managing services through explicit standards and empirical measures of performance. Operational control is assumed to rest with managers, who use performance data to monitor effectiveness, control costs, and 'hold to account' frontline practitioners. Political control is exerted in a similar way, so that managers and senior administrators are held to account for their own performance. Political and regulatory oversight of children's social care is bolstered by an independent inspectorate, Ofsted, which in its latest framework set out a 'proportionate and risk-based approach to inspection' (Ofsted, 2017). To assist with these processes, local authorities collate numerous performance data on their services for children in need, which are submitted annually to the government and published online in aggregated form (Hood et al., 2016a).

Central to PBA are two sets of distinctions, which appear in various guises in the relevant literature. The first distinction is between population-level and service-level measures. Population-level measures, which Friedman (2001) calls 'indicators', refer to an entire population as defined by geographical area or some other characteristic. An example would be the percentage of young adults in England reporting a problematic level of drug or alcohol use over the past six months (based on survey findings). Such measures attempt to reflect people's overall health and wellbeing, irrespective of whether they are receiving a service, and so are only proxies for the global attribute they describe. In contrast, service level measures relate to 'client populations', i.e. specific information held by agencies about the people they work with. According to Friedman (2001), an example of a service-level measure would be the percentage of young adults completing a drug treatment programme who have not used drugs over the past 30 days (based on self-report and keyworker assessment). Friedman argues that only this type of data should be considered as a performance measure, and that it is unfair to hold individual agencies accountable for the first category, i.e. population-wide indicators of health and wellbeing. This is a potential point of divergence from evidence-based approaches, as discussed later.

The second distinction in PBA is between processes and outcomes, which derives from the literature on quality in healthcare (e.g. Donabedian, 1966). The performance matrix developed by Friedman (1997, 2001), which is applied to child protection in Table 1, turns this binary distinction into a typology. Friedman suggests that measures of 'effort' (i.e. what services do) and measures of 'effect' (i.e. what changes) should be further subdivided into quantity and quality measures. Quantity measures summarise activity, e.g. numbers of referrals, assessments and 'case closures', while quality measures give an indication of the standard, e.g. percentage of assessments completed on time, or percentage of cases that are re-referred within twelve months of closure. Quantity of input measures (e.g. numbers of referrals or assessments) are often used to indicate levels of demand. Measures of the quality of input (e.g. timeliness of assessment completion) are sometimes referred to as 'leading' measures because they anticipate an effect on outcomes further down the line (see Pinnock, 2011); measures of the quantity and quality of effect are sometimes referred to as 'lagging' measures because they represent past performance (Parker, 2000). CP services collect a lot of lagging measures but relatively few of them represent 'quality of effect' (the bottom right quadrant in Table 1), which for Friedman are the most

useful gauge of effectiveness. For example, Hood et al. (2016b) used Friedman’s framework to categorise the performance indicators returned by English local authorities, finding only three ‘quality of effect’ measures compared to 22 other types. Moreover, two of those outcome measures were actually ‘negative’, i.e. they highlighted the proportion of children requiring repeated involvement with statutory services, which would normally be seen as undesirable (see Table 1). The third was Ofsted inspection ratings, which may be taken as an overall appraisal of performance.

Table 1. Examples of PAB measures in child protection (based on Friedman, 2001)

	<b>Quantity</b> <i>(how much did we do?)</i>	<b>Quality</b> <i>(how well did we do it?)</i>
<b>Effort</b> <i>(what did we do?)</i>	<ul style="list-style-type: none"> <li>• Number of referrals</li> <li>• Number of assessments completed</li> <li>• Number of children in need (per 10,000 child population)</li> <li>• Number of children on child protection plans</li> <li>• Number of care proceedings issued</li> </ul>	<ul style="list-style-type: none"> <li>• Percentage of assessments completed within 45 days</li> <li>• Percentage of case conferences held within 10 days of strategy meeting</li> <li>• Percentage of agency social workers</li> </ul>
<b>Effect</b> <i>(is anyone better off?)</i>	<ul style="list-style-type: none"> <li>• Number of cases closed</li> <li>• Number of care orders made</li> <li>• Number of referrals to non-statutory agencies</li> <li>• Number of referrals to specialist services</li> </ul>	<ul style="list-style-type: none"> <li>• Percentage of CP Plans where children had prior CP Plan</li> <li>• Percentage of referrals within 12 months of a prior referral</li> </ul>

Friedman’s typology raises as many questions as it answers. For one thing, child protection is a multi-agency system involved with whole families and communities. Feedback loops in such systems mean that one agency’s process measure might be seen as another agency’s outcome measure, depending on the circumstances. A recent example is the ‘KPI dataset’ assembled for the evaluation of family safeguarding teams in Hertfordshire (Forrester et al., 2017). Activity measures from health services

(children taken to hospital accident and emergency departments) and police services (adults involved in police incidents) were found to shift as a result of implementing a new approach to children's social care. These KPIs are described as a 'outcome-based performance framework' with potential applicability to children's services across the country. Of course, it may seem paradoxical to measure something as both a process and an outcome, although this is likely to occur if one agency becomes accountable for feedback loops in a wider system. Another ambiguity about Friedman's framework is the extent to which longer-term outcomes for children should be monitored. The latter has always been difficult for CP services, whose involvement with families is time-limited and circumscribed by statutory duties. While case closure may be seen as a valid marker of positive change by families and practitioners, the sustained effects of an intervention tend to be unknown unless the client re-presents to services (Barth and Reid, 2000) – in which case the intervention may be deemed to have failed! Organisational data on children's health and educational achievement are held by universal services such as schools and hospitals, and so are separated organisationally and informationally from specialist tiers and other professional domains. Theoretically it should be possible to link together these datasets, e.g. the records of children who have received statutory CP services and the National Pupil Database (DfE, 2015), although the electronic sharing of information on children is a contentious issue (Wrennall, 2010). The potential of this type of data linkage has been explored in research into outcomes for looked after children (children in public care), e.g. looking at the relationship between differential placement history and educational achievement (Maclean et al., 2017).

In the health sector, longitudinal outcome measures such as quality adjusted life years (QALY) have been widely used to monitor the success of interventions and have found a place within performance management systems, e.g. auditing, benchmarking, and payment-by-results (Smith, 2009). However, there are pitfalls to using health indicators in this way. In particular, there is the challenge of risk adjustment, i.e. identifying whether variations across groups of service users are attributable to intrinsic factors (e.g. socio-economic status), as opposed to factors that are under clinicians' or providers' control. Efforts to game the system may also emerge when performance is linked to funding and pay – a logical extension of PBA principles – leading to perverse incentives such as not treating patients with a high risk of mortality or deterioration (Smith, 2009). Indeed, there has already been controversy about payment-by-results in children's services. From 2012-16, English local authorities applied for large

amounts of government funding from the Troubled Families Programme, which was contingent on demonstrating successful outcomes for targeted families. Local authorities proceeded to make a success of the programme, leading the government to claim evidence of improvement in the lives of 99% of the families involved (Cameron, 2015). Unfortunately, these claims were later debunked by an independent evaluation, which examined a wide range of outcomes and found ‘no consistent evidence of systematic or significant impact’ (Bewley et al., 2016).

An exclusive focus on outcomes and their measurement might lead us to overlook the other side of the productivity equation, i.e. cost as well as benefit, efficiency as well as efficacy. Even assuming that valid and reliable outcome indicators could be found, the question remains as to what to use as the denominator, i.e. the inputs or resources needed to achieve a given set of results. Candidate measures might include data on expenditure and workforce, which are also collected by local authorities, or composite measures such as caseloads or ‘number of children in need per social worker’. Here too there is a problem of risk adjustment given variations in demand on resources due to local deprivation levels, for example, or demographic variables such as age and ethnicity. The conventional management approach is to vary budget allocations according to a pre-determined formula, so that budgets become the main way of monitoring and improving efficiency; indeed, another term for PBA is ‘results-based budgeting’ (Friedman, 1997). Budgets serve both as a record and a prediction for expenditure in a given domain of activity, so measures of effectiveness within that domain suggest how much value for money the service is providing. Performance data can therefore be used to inform strategic decisions on how to allocate funds between competing priorities and where to target cuts to public spending. Friedman (2001) suggests that spending on maintenance and infrastructure should be formally separated from spending which is ‘designed to produce results’, and that decisions about the latter should make use of baselines, e.g. how well a programme is working in relation to historical norms, or in relation to the performance of other comparable programmes.

## **Evidence-based approaches**

Evidence-based approaches include evidence-based practice (EBP) and evidence-based interventions (EBIs), which are fundamental to the ‘what works’ strand of evidence-based policy in health and social

care (Davies et al., 2000). In this context, evidence usually means findings from scientifically conducted research although it may include systematic knowledge produced by practitioners and service users. There is a 'hierarchy of evidence' in that some types of research have a stronger claim to validity than others. For example, an intervention that has been evaluated using a large scale randomised control trial (RCT) would be considered to be strongly evidence-based, whereas a qualitative action research project would be lower down in the hierarchy. The interface between research and practice is both direct – social workers are expected to know about research in their area of practice and incorporate it into their work – and mediated by institutional and organisational processes – social workers are expected to follow best practice guidelines and protocols for assessment and intervention that are based on research. In comparison with the PBA perspective, evidence-based approaches are therefore more ambiguous about whether control over the work lies with professionals or with managerial hierarchies.

From an evidence-based perspective, quality measures should have a clear scientific justification, i.e. a proven link between performance measures and outcomes for children. For example, La Valle et al. (2016) propose an evidence-based outcomes framework for children's social care, in which expected outcomes for children (e.g. indicators of healthy development or good-enough parenting) provide the reference point not only for measures of service quality and effectiveness, but also for the impact of social work models of intervention and the 'organisational features associated with high quality and effective services' (2016: 14). The current dearth of evidence linking together these different categories is reflected in the vagueness of policy statements on outcomes, and many process measures continue to focus on activities with limited or unproven clinical value (e.g. assessment timescales). Perhaps less obvious – because of their scarcity in child protection – is the form that process measures would have to take if they were to reflect the scientific literature more accurately. As Smith et al. (2009) point out in relation to healthcare systems, evidence that particular interventions are suited to particular populations would imply process measures with explicit inclusion/exclusion criteria, i.e. 'if...(specific population), then...(specific process)'.

Whether these criteria are understood to operate at the level of cases, interventions or services comes down to the distinction between EBP and EBIs, i.e. whether we are concerned with the 'clinical decision-making process' or with 'empirically supported interventions' (Gray et al., 2014; Thyer and Myers, 2011).

For example, a recent randomised control trial of training in motivational interviewing (MI) for child protection found it to have little effect on outcomes for families (Forrester et al., 2018). On the other hand, there is evidence that MI techniques are useful for working with people who are uncooperative or whose use of drugs or alcohol is problematic (Forrester et al., 2012). An EBP-focused process measure might therefore track the use of MI techniques in cases where parental resistance or substance misuse have been identified as a key issue; an outcome measure might track the percentage of false positives or false negatives when assessment decisions are compared to more long-term indicators, including re-referrals. In contrast, an EBI-focused process measure might track the proportion of social workers who have been trained in MI or who report themselves to be using MI in their casework. The latter would place less emphasis on ‘if...then...’ judgements in particular cases, instead assuming that as a rule social workers should be using MI in their practice. In turn, process measures would monitor the fidelity of practice to the adopted EBI; for example, services adopting motivational interviewing as an overall approach to practice might commission an external appraisal of MI practice, reporting the results as part of a performance framework (see Table 2 below). Of course, the findings from Forrester et al.’s (2018) evaluation might argue against seeing motivational interviewing as an EBI in the strict sense.

*Table 2. Examples of evidence-based measures in child protection*

	<b>Process measures</b>	<b>Outcome measures (intervention cohort)</b>	<b>Outcome measures (population)</b>
<b>Evidence-based practice (EBP)</b>	<ul style="list-style-type: none"> <li>• % of cases in which parental substance misuse is a major risk factor and an evidence-based approach used, e.g. motivational interviewing (MI)</li> </ul>	<ul style="list-style-type: none"> <li>• Self-reported or assessed shifts in parental substance misuse behaviour / readiness for change (after intervention)</li> </ul>	<ul style="list-style-type: none"> <li>• Reported levels of parental substance misuse (in the community)</li> </ul>
<b>Evidence-based intervention (EBI)</b>	<ul style="list-style-type: none"> <li>• % of CP practitioners trained in MI</li> <li>• External appraisal of MI practice</li> </ul>	<ul style="list-style-type: none"> <li>• Length of CP plans</li> <li>• Re-referral rates within 12 months</li> <li>• Repeat CP plans</li> </ul>	<ul style="list-style-type: none"> <li>• Reported levels of abuse and neglect</li> <li>• Other indicators of health and wellbeing</li> </ul>

The difference between EBP and EBIs has a bearing on whether measures are designed to promote standardisation. For example, a government commissioned report into early intervention (Allen, 2011) recommended that services should only be funded if they have been evaluated 'by at least one Randomised Control Trial (RCT) or two quasi experiments' (Featherstone et al., 2014: 68; see also Thyer and Myers, 2011). Standardisation results not only from limiting service provision to a 'suite' of evidence-based interventions but also from the need to deliver the EBI to service users in the same form every time. The basic idea is that participating in an EBI should be equivalent to receiving a dose of medical treatment; in probation services, for example, practitioners who facilitate programmes to reduce re-offending will meet regularly with a 'treatment manager', whose role is to ensure what evaluators refer to as 'programme fidelity' (Mowbray et al., 2003). By implication, organisations who are committed to delivering EBIs may wish to develop process measures for monitoring programme fidelity, including activities such as treatment management, facilitator supervision, direct observation, and external appraisal. There may even be scope for tracking longer term outcomes for people who have completed the EBI, although strictly speaking there would need to be some sort of control or comparison group for any positive changes to be considered as evidence of effectiveness. Indeed, the emphasis on scientific rigour in evidence-based approaches could deter organisations from developing their own outcome measures, leading them to concentrate instead on process measures derived from research that has been independently appraised as meeting evidential standards (Bero and Rennie, 1995; Davies and Boruch, 2001).

Unlike EBIs, the concept of EBP allows a more active role for practitioners and managers in constructing their practice and gauging its impact on service users (Thyer and Myers, 2011). Social workers are seen as more than just 'knowledge users', primed to absorb and implement the appropriate research findings; they are also 'knowledge makers' who contribute to knowledge in their field, e.g. through practice wisdom passed onto trainees and newly qualified workers (Munro, 2011). Arguably this means that practitioners still have discretion to use those interventions that seem to best suit the demands of a particular situation, and can adapt them if necessary from the evidence-based model that was originally evaluated. While this type of experimentation may not be scientifically rigorous, it may still produce useful knowledge about the mechanisms and contexts of change that can inform the 'programme theory' for a particular intervention (Pawson, 2006). However, this blurring of boundaries between

research and practice is problematic from an evidence-based perspective, which tends to emphasise the value of analytic over intuitive judgements (EBP) as well as the need for independent evaluation of social work interventions (EBIs). This division of labour between researchers and practitioners is perhaps one of the key differences between evidence-based and socio-technical approaches, as will be seen below.

A potential bridge across the research-practice divide is represented by psychometric measures, sometimes also called outcome measures or standardised instruments (see Forrester, 2017). These are tools that have been developed by researchers and/or clinicians to measure aspects of someone's health, wellbeing, attitudes or behaviour. They often take the form of questionnaires about how an individual feels or functions, and are generally filled in by the person in question, a family member or professional. One example familiar to many CP social workers is the Strengths and Difficulties Questionnaire (Goodman, 1997), which was part of a 'family pack of scales and questionnaires' released by the UK government to accompany a statutory framework for child and family assessments (Cox and Bentovim, 2000). Such questionnaires are considered to be evidence-based because their psychometric properties, i.e. validity and reliability, have been scientifically tested. Some measures can be used to identify clinical 'cut-offs' based on the comparison with population norms, which may contribute to decisions about thresholds of harm or parenting capacity. In child protection, psychometric measures tend to be the province of psychologists rather than social workers, although social workers conducting parenting assessments will also be familiar with them. Shifts in outcome measures over time can demonstrate change (positive or negative) in particular characteristics and may therefore confirm or refute assessments based on professional judgement and qualitative information. In principle, measures that are consistently used by practitioners could also be aggregated into service-level indicators, tracking both patterns of need and the impact of interventions.

## **Socio-technical systems design (STSD)**

Socio-technical systems design (STSD) originated with the work of Erik Trist and colleagues at the Tavistock Institute of Human Relations in the late 1940s and 1950s (Trist, 1981). The underlying premise is that systems involve a complex interaction between people, technology, and other factors in the environment. General systems ideas such as interdependence and equifinality are developed into design

principles such as multifunctionality and minimal critical specification (Mumford, 2006). STSD also adopts a humanist perspective on work, e.g. it is considered important that people's jobs are intrinsically motivating and allow for personal and professional development (rather than being dull, repetitive and procedural). It represents a critique of Taylorist management techniques and technocratic assumptions about perfectible processes, inclining towards decentralised organisational structures and the shifting of operational control near to frontline practice rather than at the top of bureaucratic hierarchies. It has found applications in a wide range of settings and sectors, including coal mining, car manufacturing, air traffic control and nuclear energy (Woods et al., 2010).

Eileen Munro drew on socio-technical ideas in her analysis of the UK child protection system, arguing that over-prescription and proceduralisation of CP work was eroding the professional expertise needed to work effectively with children and families (Munro, 2010). Her review recommended that organisations change the way they used performance information, including feedback from inspections and serious case reviews, in order to increase expertise and capability in frontline practice. For example, while the principle of 'timeliness' was important in CP work, it had to be seen from the perspective of the child rather than that of the agency. Munro's final report therefore recommended the removal of prescribed timescales for individual assessments, seeing them as counterproductive because they focused on organisational tasks rather than what was important to families (Munro, 2011). Other measures of timeliness were seen as potentially more useful, e.g. the time taken for effective help to be provided (including repeated referrals and assessments), or the number of handovers to other social workers and professionals.

Another example of a socio-technical approach is John Seddon's Vanguard Method (Seddon, 2003), which has been used in the UK to redesign child protection (Gibson and O'Donovan, 2014) as well as adult social care (e.g. Newmann and Jones, 2016). Seddon argues that good measures should reflect the purpose of a service from the end-user's point of view, demonstrate variation over time, and be a tool for practitioners doing the work rather than used as an arbitrary standard or target. Operational flexibility is essential so that practitioners can respond to the variety of demand, and so rather than using performance data to appraise and control the work of subordinates, managers should work alongside practitioners to understand and improve how the system works. Seddon distinguishes

between 'value demand', which represents what citizens want from a service, and 'failure demand', which results from services not dealing properly with value demand. This distinction is not captured in conventional process measures, such as numbers of referrals, assessments or CP plans, which take no account of how long children have had to wait or be re-referred until they receive an effective form of support or intervention. For example, Gibson and O'Donovan's (2014) study found that that failure demand in CP services was running at up to 70% of total demand and average end-to-end time from the child's perspective (i.e. counting re-referrals) was a shocking 852 days. System conditions responsible for this state of affairs included an inflexible electronic workflow system, preoccupation with gatekeeping, managing demand and cutting costs, use of numerical targets to manage performance, fragmented provision with multiple transitions between teams, and services not addressing the underlying causes of problems. Following redesign, failure demand was reduced to about 25-30%, decision-making was decentralised with managers spending much more time working alongside practitioners, social workers spent 80-90% of their time out face-to-face with the people they were supposed to be helping, and end-to-end times were substantially reduced.

One challenge of STSD in people-centred services is how to align individual measures – have we done what matters to this person? – with system measures – how have we done overall? In evidence-based approaches, this issue is resolved by linking clinical guidelines to appropriate process measures, e.g. the percentage of service users with a particular problem who are receiving the recommended treatment or intervention, and these are linked in turn to outcome measures. From a socio-technical perspective, the drawback of tailoring measures to the scientific literature, rather than to the full range of problems and solutions encountered in practice, is that resources and provision tend to converge on a 'one size fits all' template that can frustrate people's efforts to get what they really need. This is a particular issue for social care services, which try to control demand through lengthy assessment forms composed of a-priori information categories. The psychometric measures discussed earlier are also inflexible in many respects, e.g. in the order and wording of questions. An alternative approach is to accept more ideographic types of evidence, e.g. narrative pen-portraits or self-report scaling tools, in order to understand the issues that service users consider to be important. For example, Forrester (2017) reports the use of user-report tools such as Goal Attainment Scaling (Kiresuk and Sherman, 1968) and the Satisfaction with Life Scale (Diener et al., 1985) in order to evaluate child safeguarding interventions.

Such tools are less reliable when it comes to making comparisons over time and between people, but are personalised in a way that psychometric measures and assessment forms are not. Indeed, in Wales the code of practice accompanying the 2014 Social Services and Well-being (Wales) Act states explicitly that 'local authorities need to understand what matters to people and the personal outcomes they wish to achieve' (Llywodryeth Cymru Welsh Government, 2016: 8).

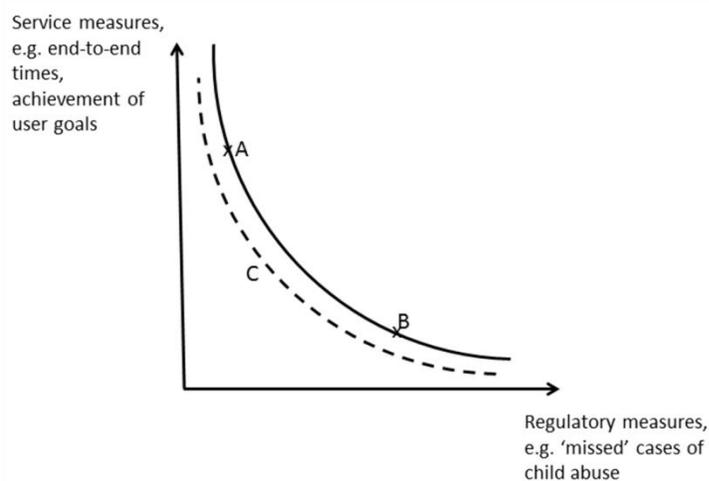
Table 3. Examples of measures in socio-technical systems design

	Qualitative	Quantitative
<b>Individual measures</b>	<ul style="list-style-type: none"> <li>• Pen-portraits</li> <li>• ‘What matters’ conversations</li> <li>• Outcome star/spidergram</li> </ul>	<ul style="list-style-type: none"> <li>• Scaling (e.g. GAS, Life Scaling, spidergram)</li> <li>• Self-report measures (including psychometrics)</li> </ul>
<b>System measures</b>	<ul style="list-style-type: none"> <li>• Analysis of system conditions, e.g. single and double feedback loops</li> </ul>	<ul style="list-style-type: none"> <li>• End-to-end times (from first contact to problem solved)</li> <li>• Number of handovers</li> <li>• Rate of failure demand</li> <li>• Regulatory measures, e.g. ‘missed’ abuse cases</li> </ul>

Individual and system measures are not always well aligned in child protection. For example, it is possible for the latter to register ‘success’ (e.g. an assessment completed on time and appropriate referral made) when the former would register ‘failure’ (e.g. the citizen does not engage with drug and alcohol services and is evicted from her property). In part, this may be attributed to the influence of performance management frameworks. However, a more fundamental problem is that the primary ‘users’ of services, i.e. children and their families, must be considered alongside other stakeholders, including the general public, which funds services through taxation. Defining the overall purpose of CP in order to derive socio-technical measures must therefore accommodate competing interests and goals. One way to conceptualise this is by separating the regulatory role of agencies, which is about detecting, assessing and intervening in cases of abuse, from their service provider role, which is more generally about promoting the welfare of children. This separation is already implicit in the way services are organised to reflect their legislative underpinning (see Hood, 2015), even if the idea of ‘partnership working’ is intended to resolve the potential tension between ‘care’ and ‘control’. As Sparrow (2000) points out, regulatory problems are not always amenable to process improvement, partly because ‘demand’ for services does not capture harm that is invisible to – or being kept hidden from – existing referral mechanisms. Moreover, services designed to ensure ‘citizen satisfaction’ are not necessarily effective from a regulatory point of view. Sparrow (2000: 67) calls this a ‘facilitation-enforcement trade-off’, an adapted version of which is illustrated in Figure 1. In CP, this is sometimes called the ‘false

positives' problem because pressure on agencies not to miss potential child abuse leads to higher numbers of inappropriate investigations. Shifting to a less investigatory regime, represented as a shift from point B to point A in Figure 1, might make services less oppressive and more supportive but could also increase the number of 'false negatives', i.e. unidentified cases of abuse. Efforts to improve performance on both counts would involve shifting to a curve closer to the origin, represented by curve C in Figure 1, but would not eliminate the trade-off itself.

Figure 1. Trade-off between regulatory and service measures (adapted from Sparrow, 2000: 67)



The notion of regulatory trade-offs illustrates the importance of knowing how measures relate to each other and with the system as a whole. In another example, Hood et al. (2016a) found that local authorities in more deprived areas experienced higher overall levels of demand for children's social care, tended to screen out more referrals, close down cases more quickly, were less likely to work longer term with families, and also had higher rates of re-referral within 12 months. However, making the link between 'leading' and 'lagging' indicators in this way requires the disentangling of complex causal relationships. In evidence-based approaches, measures representing causal relationships are used to operationalise a 'theory of change' or 'programme theory'. For example, a parenting programme might have been found to help participants to regain a sense of self-efficacy, act more calmly in challenging situations, and empathise with their children's reasons for behaving in particular ways (Barlow and

Stewart-Brown, 2001). A local authority might then commission this programme as an EBI, incorporating it into the social care services for children and families deemed to be 'in need'. However, measures used to evaluate one programme's effectiveness may not be good proxies for the effectiveness of children's social care as a whole. The latter requires measures that are tailored to overall purpose, rather than to the EBI in part of the system.

In relation to cost – the 'efficiency' part of the productivity equation – STSD raises some interesting questions about the role of budgets and unit costs in planning and administration. Seddon (2003) shows that public services often seek to reduce the cost of individual transactions and activities, e.g. by outsourcing referrals to a call centre and paying as little as possible 'per call'. Often the unintended consequence is a fragmented and ineffective service, leading to higher levels of failure demand and higher costs across the system. The implication for child protection is that higher unit costs at the point of referral, which tend to result from moving expertise and resources to the frontline, will be more than compensated for by the speed with which value demand is addressed and failure demand is reduced (see Gibson and O'Donovan, 2014). A similar critique can be applied to budgets, which according to Bogsnes (2016) provide a floor to costs as much a ceiling and focus managerial attention on 'hitting the numbers' rather than on studying demand and experimenting with design. That is not to say that expenditure should not be measured and monitored within a STSD but rather that the work of different subsystems (e.g. functions, professional groups) should be focused on the system's overall purpose rather than on meeting internal specifications. A further implication is that process measures oriented towards discrete tasks and activities, such as assessment timescales, may end up distorting practice to the detriment of the overall service.

## **Conclusion**

This paper began by asking a simple question (what to measure in child protection?) and in typically academic fashion has proceeded to point out how complicated it is. An ostensibly more satisfying response might have been to list all possible measures within a unifying framework, which is what tends to happen within each of the perspectives discussed above. Instead, the concern here has been to discuss the perspectives themselves and point out some inconsistencies and contradictions between

them. Of course, their differences are arguably outweighed by a common empiricist assumption, namely that it is possible to quantify what effective (and efficient) child protection looks like. Having made this assumption, each perspective makes a case for its own version of what should be quantified, e.g. quality in accountability frameworks, outcomes in evidence-based approaches, or purpose in socio-technical systems design. Assumptions about the institutional context of child protection vary, so that practitioners could be operating at the bottom of managerial hierarchies or within decentralised work groups. Accountability approaches suggest that it is unfair to assess the impact of services on long-term indicators of health and wellbeing; evidence-based approaches insist that it is essential to do so. These contextual factors provide each perspective with its internal logic, a consensus on what should be measured and why it makes sense to measure it. In the process, the socio-political backdrop to consensus is often forgotten, leaving a reified concept: the data that speaks for itself.

Like any other public service, child protection serves multiple stakeholders: children, first and foremost, but also parents, communities, taxpayers, politicians, inspectors, and so on. Each of these groups will understand effectiveness and efficiency – and the measures associated with those concepts – in different ways. In political and professional discourse, these differences are usually elided by insisting on the primacy of outcomes for the child. However, as this overview has tried to show, the nature of outcomes and their link to service characteristics is a complex and disputed territory. Moreover, the interests of stakeholders in child protection are not always perfectly aligned. The sector's political sensitivity makes it reasonable for politicians and managers to equate measures with standards, and to enforce those standards through targets, inspection regimes, quality assurance processes and the like. Whether the interests of children and families are best served by such strategies is debatable. An adaptation of Goodhart's Law might be: when a measure is used for accountability purposes it stops being a good measure of anything else!

While it is easy to criticise the use of measures within a specification and compliance culture, it is important not to overlook the pitfalls with other approaches to measurement. Evidence-based approaches are also prescriptive in some respects, e.g. the hierarchy of evidence, the separation of research and evaluation from practice and management, and the top-down implementation of EBIs. Measures deriving from such approaches may simply be incorporated within existing performance

management frameworks, particularly in an institutional context characterised by centralised organisational structures and managerial hierarchies. The approach to measurement in socio-technical systems design represents perhaps the most radical break with current practices, particularly in England, and so is arguably the least likely to be adopted. Indeed, perhaps the main criticism of STSD in the current climate of child protection is that it appears to present an idealistic rather than a pragmatic solution to the dilemmas faced by the sector. But then again: maybe it is time to give it a try.

## References

- Barlow, J. and Stewart-Brown, S. (2001) Understanding parenting programmes: parents' views, *Primary Health Care Research and Development*, 2(2), pp. 117-130.
- Bero, L. and Rennie, D. (1995) The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Jama* 274(24), pp.1935-1938.
- Bewley, H., George, A., Rienzo, C. and Porte, J. (2016) *National Evaluation of the Troubled Families Programme: National Impact Study Report: Findings from the Analysis of National Administrative Data and local data on programme participation*, London, Department for Communities and Local Government.
- Bilson, A. and Martin, K. (2016) Referrals and Child Protection in England: One in Five Children Referred to Children's Services and One in Nineteen Investigated before the Age of Five, *British Journal of Social Work*, 47(3), pp. 793-811.
- Bogsnes B. (2016) *Implementing beyond budgeting: unlocking the performance potential*, Hoboken, NJ, John Wiley and Sons.
- Bywaters, P., Brady G., Sparks, T., Bos, E., Bunting, L., Daniel, B., Featherstone, B., Morris, K. and Scourfield, J. (2015) 'Exploring inequities in child welfare and child protection services: Explaining the 'inverse intervention law'', *Children and Youth Services Review*, 57, pp. 98-105.
- Cameron, D. (2015) Press release: PM praises Troubled Families programme success. Available online: <https://www.gov.uk/government/news/pm-praises-troubled-families-programme-success>
- Cox, A. and Bentovim A. (2000) *Framework for the Assessment of Children in Need and their Families: The Family Pack of Questionnaires and Scales*, London, The Stationery Office.
- Davies, H., Nutley, S. and Smith, P. (2000) *What works? Evidence-based policy and practice in public services*, Bristol, Policy Press.
- Davies, P. and Boruch, R. (2001) 'The Campbell Collaboration: does for public policy what Cochrane does for health', *British Medical Journal*, 323, pp. 294-295.
- Department for Education (DfE) (2015). *Children's safeguarding performance information framework*. Available online, <https://www.gov.uk/government/publications/childrens-safeguarding-performance-information-framework>.
- Diener, E., Emmons, R.A., Larsen, R.J. and Griffin, S. (1985) 'The satisfaction with life scale', *Journal of personality assessment*, 49(1), pp. 71-75.
- Donabedian, A. (1966) Evaluating the quality of medical care. *The Milbank memorial fund quarterly* 44: 166-206.
- Forrester, D., Westlake, D. and Glynn, G. (2012) Parental resistance and social worker skills: Towards a theory of motivational social work. *Child and Family Social Work* 17: 118-129.
- Forrester, D., Westlake, D., Killian, M., Antonopoulou, V., McCann, M., Thurnham, A., Thomas, R., Waits, C., Whittaker, C. and Hutchison, D. (2018) 'A randomized controlled trial of training in Motivational Interviewing for child protection', *Children and Youth Services Review*, 88, pp. 180-190.
- Forrester, D. (2017) 'Outcomes in Children's Social Care', *Journal of Children's Services*, 12(2/3), pp. 144-157.
- Freidson, E. (2001) *Professionalism: The Third Logic*, Cambridge, Polity Press.
- Friedman, M. (1997) *A Guide to Developing and Using Performance Measures in Results-Based Budgeting*, Washington, DC, Finance Project.

- Friedman, M. (2001) *The results and performance accountability implementation guide*. Available online at: [www.raguide.org](http://www.raguide.org).
- Gibson, J. and O'Donovan, B. (2014) 'The Vanguard Method as applied to the design and management of English and Welsh Children's Services departments', *Systemic Practice and Action Research*, 27, pp. 39-55.
- Goodman, R. (1997) 'The Strengths and Difficulties Questionnaire: A Research Note', *Journal of Child Psychology and Psychiatry*, 38, pp. 581-586.
- Gray, M., Joy, E., Plath, D. and Webb, S.A. (2014) 'Opinions about evidence: A study of social workers' attitudes towards evidence-based practice'. *Journal of Social Work*, 14, pp. 23-40.
- Hood, C., Rothstein, H. and Baldwin, R. (2001) *The government of risk: Understanding risk regulation regimes*, Oxford, Oxford University Press.
- Hood, R., Goldacre, A., Grant, R. and Jones, R. (2016a) 'Exploring demand and provision in English child protection services', *The British Journal of Social Work*, 46(4), pp. 923-941.
- Hood, R., Grant, R., Jones, R. and Goldacre, A. (2016b). A study of performance indicators and Ofsted ratings in English child protection services, *Children and Youth Services Review*, 67, pp. 50-56.
- Hood, R. (2015). A socio-technical critique of tiered services: Implications for interprofessional care, *Journal of Interprofessional Care*, 29(1), pp. 8-12.
- Kiresuk, T. J. and Sherman, R.E. (1968) 'Goal attainment scaling: A general method for evaluating comprehensive community mental health programs', *Community Mental Health Journal*, 4(6), pp. 443-453.
- La Valle, I., Holmes, L., Gill, C., Brown, R., Hart, D. and Barnard, M. (2016) *Improving Children's Social Care Services: Results of a feasibility study*, London, CAMHS Press.
- Levy, R. (2010) 'New Public Management', *Public Policy and Administration*, 25, pp. 234-240.
- Llywodryeth Cymru Welsh Government. (2016) *Social Services and Well-being (Wales) Act 2014. Part 2 Code of Practice (General Functions)*. Available online at: <http://gov.wales/docs/dhss/publications/151218part2en.pdf>.
- Maclean, M.J., Taylor, C.L. and O'Donnell, M. (2017) 'Relationship between out-of-home care placement history characteristics and educational achievement: A population level linked data study', *Child Abuse and Neglect*, 70, pp. 146-159.
- Mumford E. (2006) 'The story of socio-technical design: Reflections on its successes, failures and potential', *Information Systems Journal*, 16, pp. 317-342.
- Munro, E. (2004) 'The impact of audit on social work practice', *British Journal of Social Work*, 34, pp. 1075-1095.
- Munro, E. (2010) *The Munro Review of Child Protection Part One: A Systems Analysis*, London, Department for Education.
- Munro, E. (2011) *The Munro Review of Child Protection: Final Report. A child-centred system*, London, The Stationery Office.
- Newmann, J. and Jones, S. (2016) *Report to Healthier Communities Committee: Vanguard - Progress Made in Implementing the New Way of Working in the North Norfaen Wellbeing Team (NTWBT)*. Available online at: <http://modern.gov.torfaen.gov.uk/documents/g2853/Public%20reports%20pack%2017th-Nov-2016%2010.00%20Healthier%20Communities%20Overview%20Scrutiny%20Committee.pdf?T=10>.

- Office for National Statistics (ONS) (2016) *Productivity Handbook*, Available online: <https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/methodologies/productivityhandbook/pdf>.
- Parker, C. (2000) 'Performance measurement', *Work study*, 49(2), pp. 63-66.
- Pawson, R. (2006) *Evidence-based Policy: A Realist Perspective*, London, Sage.
- Pinnock, M. (2011) *Canary in the Cage? Lead Indicators and their potential use by Local Safeguarding Children Boards and partner agencies*, London, Childhood Wellbeing Research Centre.
- Schorr, L. (1995) *The Case for Shifting to Results-Based Accountability with a Start-Up List of Outcome Measures*, Washington, DC, Centre for the Study of Social Policy.
- Seddon, J. (2003) *Freedom from command and control*, Buckingham, Vanguard Education.
- Shaw, I., Bell, M., Sinclair, I. (2009) An Exemplary Scheme? An Evaluation of the Integrated Children's System. *British Journal of Social Work*, 39(4), pp. 613-626.
- Smith, P. C. (2009) *Performance measurement for health system improvement: experiences, challenges and prospects*, Cambridge, Cambridge University Press.
- Sparrow, M.K. (2000) *The regulatory craft: controlling risks, solving problems, and managing compliance*, Washington, DC, Brookings Institution Press.
- Trist, E. (1981) *The evolution of socio-technical systems: a conceptual framework and an action research programme*, Toronto, Ontario Quality of Working Life Centre.
- Thyer, B. A. and Myers, L. L. (2011). The quest for evidence-based practice: A view from the United States. *Journal of Social Work*, 11(1), pp. 8-25.
- Van Dooren, W., Bouckaert, G. and Halligan, J. (2015) *Performance management in the public sector*, Abingdon, Routledge.
- Woods, D., Dekker, S. and Cook, R., (2010) *Behind Human Error*, Farnham, Ashgate.
- Wrennall, L. (2010) Surveillance and child protection: De-mystifying the Trojan Horse. *Surveillance and Society*, 7, pp. 304-324.