­­

**The Fifth Phase of Educational Effectiveness Research: the philosophy and measurement of equity**

Anthony Kelly

*“The whole point of justice consists in our providing for others through humanity what we provide for our own family though affection.”* Lactantius (4th Century philosopher).

**Introduction**

Educational Effectiveness Research (EER) investigates, and seeks to explain, the causal relationships in formal school settings between on the one hand *outcomes* (both academic-cognitive and educational-affective) and on the other, *inputs* (student intake characteristics such as social background and prior attainment). Of course, ratios of output-to-input are simply measures of *efficiency*, not effectiveness, but EER goes beyond this to look at the educative *processes* between inputs and outputs[[1]](#footnote-1) and the *policy context* in which these processes are set, and this is what justifies ‘effectiveness’ in the nomenclature. Recently there has been a certain nervousness within the field about the formality of the institutional setting within the EER paradigm - usually schools but more recently pre-schools and colleges of further education – and this has manifested itself in a desire to be more inclusive in how the field is defined, as if the mere fact of blurring the boundaries of the discipline will meet some higher social aspiration. The practical reality remains that usually only *formal* systems collect the kind of empirical data that underpins EER methodologies and enables modelling to take place at the level of the pupil, the teacher, the school and the system, although some EER studies have also added measures of out-of-school learning (e.g. engagement in private tutoring, time spent on homework and other home learning environment measures) to provide additional controls of intake differences between schools or cohorts and to provide insights into the potential importance of such out-of-school learning opportunities.

The educative processes (broadly defined) within EER constitute the sub-field of School Improvement Research (SIR), which investigates factors such as classroom teaching, curriculum, school and classroom behaviour, learning climate, organisation and leadership, and their effect on student outcomes / outputs. These processes are closely linked to policy and system-level factors, just as inputs are intrinsically linked to the social and cultural milieu of the school, but in its forty-year-old morphology, EER has tended to overlook the impact of context on *outcomes*: the field measures *what society deems valuable*. This is not ‘context’ in the usual sense, but it is a context nevertheless - a *philosophical* one – and it is something that should not be ignored, especially as EER moves into a new phase where it is applied to the issue of equity.

**How the phases of EER have developed**

EER accepts that differences in pupil attainment are largely determined by ‘ability’, however that is defined, and by socio-economic and other factors linked to family, school and neighbourhood, but that schools can and do make a significant difference. It attempts to explain why, how and to what extent, some schools and teachers are more effective than others in achieving desirable outcomes for pupils. It is a quantitative, institutional-focused approach that began in earnest in 1979 as a critique of earlier US research on equity and opportunity carried out by Coleman et al. (1966), who found that only an *insignificant* amount of variance in pupil outcomes was explained statistically by school-level factors and that after taking into account pupil background characteristics such as ability and socio-economic status, only a small proportion of the variation in pupil attainment could be attributed to differences between individual schools. EER was born of a desire to counteract this pessimism in an era that was marked by social frivolity and the dysfunction of schooling for socially disadvantaged pupils, and of not knowing what, if anything, we could do to about it.

The first studies, undertaken by Edmonds (1979) in the US and Rutter, Maughan, Mortimore and Ouston (1979) in the UK, showed that schools *could* make a difference. Both studies used similar quantitative methodologies in looking at effectiveness between (rather than within) schools. The unit of assessment was the institution, the approach was scientific and the findings (like the initial motivation) were optimistic. EER became the thing it wanted to be, and as time passed and more sophisticated approaches to data modelling were developed, it drove the science upon which many national policies and international comparisons were based. Put simply, it involved measuring a school’s output in terms of pupil attainment, correcting for input, circumstance and context, and assigning a scalar to the school effect.[[2]](#footnote-2)

Creemers and Kyriakides (n.d.) have identified four historical phases in the development of the field since 1979, each addressing different research questions and adopting different theoretical stances (see Figure 1). Phase One, from 1979 to the mid-1980s, focused on the size of the school effect, showing (as it was designed to show) that the effect was big enough to matter. Studies looked at differences in the impact that particular teachers and schools had on pupil attainment (ignoring possible confirmation bias) and finding that school and teacher effects were larger for disadvantaged groups (Scheerens & Bosker, 1997). Phase Two, from the mid-1980s to the early 1990s, focussed on the characteristics of effective schools and the factors associated with good teaching and better pupil outputs. In the manner of leadership research within SIR, it produced lists of factors (and lists of lists) characteristic of effective teachers and schools. Phase Three, from the 1990s to the early 2000s, developed theoretical models to explain why some factors, but not others, were important in explaining variations in pupil attainment. Various integrated models of educational effectiveness were developed (Creemers, 1994; Scheerens, 1992; Stringfield & Slavin, 1992) and according to Creemers and Kyriakides (n.d.) these ‘guided not only the theoretical development of EER but also the design of empirical studies in the field’ (Kyriakides et al., 2000; de Jong et al., 2004).

Phase Four, from the mid-2000s to the present day, seeks to take better account of context and re-engage with the processes of SIR, especially the kind of improvement that is sustained over time. It acknowledges that effectiveness is not a *stable* trait of schools or classrooms, but is ‘dynamic and complex’ (Creemers & Kyriakides, 2008), that it can vary from pupil to pupil and can be differentially effective in respect of different outputs. Historians of EER claim that this fourth phase places ‘change’ at the heart of the field, but this is perhaps to overstate the metamorphosis. Change is merely the absence of certainty and is to EER what the equator is to geography; the *product* of a methodology rather than an object in and of itself. Nevertheless, Phase Four has moved EER into an era of profitable, large-scale, systematic evaluations of policy effects where it has formed an intellectual alliance with transnational organisations like the Organisation for Economic Cooperation and Development (OECD). The dangers of such associations have perhaps been downplayed - as Chekov said in *The Cherry Orchard* “If you run with the hounds you are entitled to bark, but you are also expected to wag your tail’ - but the problem of ‘context’ continues as a motif across the field as context at the level of the school is replaced by context at the level of the nation state, without either being fully satisfied.

The unsatisfactory nature of how context is accommodated methodologically and philosophically is a stone-in-the-shoe of EER, particularly when it comes to equity and fairness, an area where ‘context’ (in every sense of the word) is critical. This suggests that EER will need to move into a new fifth phase of development as it shifts its main focus to effectiveness-for-equity (see Figure 1). This will require a fundamental consideration of the field’s underpinning philosophy. EER has a long tradition of addressing the attainment gap between the ‘haves’ and the ‘have-nots’ (Sammons, 2007) and it has always looked at variations in practice between schools and between classrooms with a view to making educational outcomes for disadvantaged pupils more equitable and fairer. The challenge for Phase Five will be to debate and define what is meant by ‘fairness’ and ‘equity’ and in what circumstances a degree of inequality is acceptable. Phase Five is akin in some ways to the way theoretical physics moved a century ago from the methodology of the laboratory to the metaphysical paradigm of quantum physics, where previously accepted certainties like time and position had to be defined. So it is with EER. Like Dirac’s early work on quantum theory where there were too many ‘infinities’ (Kragh, 1990), EER has left too much undefined since 1979 in the field’s rush to action. What do we mean by effectiveness and equity? How much of what we (think we) know is itself the product of how it has been investigated? And when we talk about education reducing inequality in society, who do we have in mind and why? To date, EER has not articulated a coherent philosophical underpinning (Kelly & Elliott-Kelly, 2018). This may be the result of paying so much attention to methodology over the first four phases of its development and to the rich datasets available to which that methodology has been applied. Phase Five will define generally what is meant by ‘equity’ in terms of justice and fairness; specifically, to address the problem that schools can improve dynamically in aggregate but adversely affect disadvantaged pupils, and that more effective schooling can sometimes exacerbate achievement differentials in society. This chapter proposes a philosophical underpinning for equity and goes on to look at its measurement.[[3]](#footnote-3) It suggests that the utilitarianism into which EER defaulted *faute de mieux* is no longer appropriate for the moral imperatives at the heart of the paradigm, and that an alternative philosophy based on Rawls’s theory of ‘justice as fairness’ is one way forward, especially in progressing the dynamic theoretical model of Phase Four (Creemers & Kyriakides, 2006 & 2008), which treats equity, measured crudely (and wrongly, as an output) by the size of the attainment gap between disadvantaged and advantaged pupils, as a critical internal component.

INPUTS

e.g. prior attainment

PROCESSES

e.g. teaching & school processes

OUTPUTS

e.g. pupil attainment

OUTCOMES

e.g. equity & a fairer society

*Context:*

*policy*

*Context:*

*socio-cultural*

*Context:*

*policy*

*Context:*

*philosophical*

EER Phases 1-4

EER Phase 5

Figure 1: Phases of EER development

**PART 1: The philosophy of equity**

**EER and a creeping utilitarianism**

Utilitarianism is a philosophy that developed over the course of the Eighteenth Century. In its original form, as developed in 1726 by Francis Hutcheson and more famously in 1780 by Jeremy Bentham, it holds that the best action is the one that maximises utility for the greatest number of people; that is to say, that virtue is in proportion to the number of people that derive benefit from it, and that the best course of action is therefore the one that secures the greatest happiness for the greatest number. Utility is defined as the sum of all the benefit, minus the detriment, that results from an action. In an educational setting, we can equate it with attainment and achievement.

The two defining characteristics of utilitarianism are the aggregation (or averaging) of benefit, and a reliance on the measurement of proxy outcomes. In utilitarianism, as with EER and SIR, actions are judged by their efficacy, convenience and advantage. The outcome of any action is the sole measure of whether it is right or wrong. Utilitarianism comes in several forms determined by their preferred proxy methods for measuring utility, like EER and SIR. Total Utilitarianism, for example, calculates utility as an aggregate, whereas Average Utilitarianism calculates it as an average. Henry Sidgwick, in his 1874 book *The Methods of Ethics*,[[4]](#footnote-4)discusses the question of *aggregate* versus *average* utility, concluding that what should be maximised is the average utility multiplied by the number of people in the population. This means that if average utility remains constant, utilitarianism demands that we make the number of people benefitting as great as possible. Both aggregate and average utility might seem uncontroversial, but problems of moral alignment emerge when we apply either to education and schooling. Maximising *average* utility can mean ignoring all pupils whose educative benefit is below average, which might not be what a community wants at any given time; and maximising *aggregated* utility can lead to the situation where large numbers of pupils having very small educative benefit is regarded as a better outcome than a smaller number of pupils having larger benefit, which again might not be what a society needs (economically or socially) at any given time.

Bentham (1780) stated the Principle of Utility as the principle that ‘approves or disapproves of every action according to the tendency’ it has to ‘augment or diminish happiness’. In his so-called ‘hedonic calculus’, he suggested that the value of a pleasure is measured by the number of people affected. Hutcheson eventually rejected this algorithm as ‘useless and disagreeable’, but Bentham could see ‘nothing unwarranted’ in it. Philosophically, this echoes the fixation among policy makers in education with measuring the achievement of intangibles, or failing that, concentrating on what is measurable and ignoring other factors that are important but difficult to gauge (Kelly & Elliott-Kelly, 2018). Despite the advocacy of Creemers and Kyriakides (2008) whose Phase Four dynamic model proposes including non-traditional metrics like well-being (see also Opdenakker & Van Damme, 2000), EER has, over the course of its four phases of development, been driven (or at least, encouraged) unwittingly into its own spurious calculus by policy-makers for whom a utilitarian view of schooling is convenient. This imperative, officially sanctioned, has been secured through targeted funding from quasi-government sources like research councils, and the privileging of certain lines of research inquiry and certain econometric methods at the expense of other approaches.

Bentham’s view that in promoting greater utility, ‘governments should punish’ in proportion to the extent to which certain actions are ‘pernicious’[[5]](#footnote-5) also has echoes in today’s education policy landscape of accountability and naming-and-shaming schools that do not come up to some notional aggregate water mark. This is not the fault of EER *per se* – the field cannot be blamed for the uses to which its findings are put - but it does show how well-aligned the field has become with the utilitarianism of western government policy and that of transnational bodies like the OECD. *'Natural justice demands that schools are held accountable only for those things they can influence (for good or ill) and not for all the existing differences between their intakes’* (Nuttall, 1990, p 25), so for EER, exploring the impact of different intake factors is seen as crucial to the field’s attempt to promote social inclusion and widen the social distribution of achievement. There was early recognition by many in EER that there were limits to how much schools could compensate for / ameliorate wider disadvantage in society (Mortimore et al, 1988; Mortimore & Whitty, 1997), but in the absence of any clearly formulated and asserted philosophy to the contrary, policy-makers unhampered by any pedantic adherence to the evidence have steered some EER (especially by bodies such as the OECD and TIMSS) towards an outlook where utilitarianism seems to have become the default paradigm (Kelly & Elliott-Kelly, 2018).

Modern utilitarianism began with John Stuart Mill. He was a follower and promoter of Bentham’s ideas, but his 1863 book *Utilitarianism* rejected as absurd the latter’s purely quantitative measurement of utility. Although Mill asserted the ‘pleasures of the intellect’ over the more inclusive hedonism of Bentham, the two were agreed that the welfare of the majority was always to be paramount. In Chapter 4 of his book, Mill offered his famous proof for the Principle of Utility: that greater utility is desirable and is ‘a good to the aggregate of all persons’ because people make it so, in the same way that the proof that an object is visible is that people can see it. This ‘notorious’ (as Alican, 1994, called it) argument is fallacious on several counts. Firstly, Mill is inferring what people *ought* to do from what they *actually* do; a naturalistic fallacy. Secondly, he is inferring that something *ought* to be desired from the fact that it is *capable* of being desired; a fallacy of equivocation. And thirdly he is inferring that because people desire greater utility *for themselves*, that the aggregate of all persons will desire greater utility *generally*, and that this will be the *only* thing they desire.

The Twentieth Century saw the development of new types of utilitarianism: Act Utilitarianism and Rule Utilitarianism, both of which emphasise the central role of regulation in helping people chose the course of action that maximises utility. The difference between Act and Rule lies in how the action is judged to be the right one: Act Utilitarianism holds that an action is right if *that action* maximises utility; Rule Utilitarianism maintains that an action is right if it conforms *to a rule that maximises utility*. The difference is really about the general versus the specific, and Hare (1973 & 1981) developed this into a two-level theory to bring forth the distinction between the two:

* Specific-rule Utilitarianism, which Hare suggests we use when we are deciding what principles to follow. Specific-rule Utilitarianism reduces to Act Utilitarianism.
* General-rule Utilitarianism, which Hare suggests we use this when we are in a situation where natural bias is likely to prevent us from calculating the best course of action. General-rule Utilitarianism does *not* collapse into Act Utilitarianism.

Hare (1981) illustrates his two-level theory using a thought experiment similar to that of John Rawls (1971) a decade earlier, which we will discuss in the next section. Hare conjures up two archetypes to represent the two extremes of ‘general’ and ‘specific’. The Archangel is a hypothetical person who has perfect knowledge of every situation, has no personal bias and always uses critical thinking to do the right thing. The Prole, on the other hand, is completely incapable of critical thinking, uses only intuition and from necessity blindly follows general rules. Hare is not suggesting that people are either Archangels or Proles, but rather that everyone has the characteristics of both to varying degrees, in different contexts and at different times (Kelly & Elliott-Kelly, 2018). Hare did not specify when and under what conditions people act as Archangels and when and under what conditions they act as Proles, but his ‘dramatic device’ is important in illustrating the primacy in Hare’s utilitarianism of formal *critical* *thinking* (the Archangel’s modus operandi) over experiential *intuition* (the Prole’s modus operandi). This has parallels in the development of EER, which has come to regard its own modeling as having primacy over the intuition of teachers. In fact, this is acknowledged implicitly in the successful dynamic model of Creemers and Kyriakides (2008). It is quite proper that autonomous practitioners in any profession should default to critical thinking when working in unusual situations, but in the context of professional practice, whether in schools by teachers or in hospitals by doctors, it is a mistake to regard professional intuition as being devoid of criticality. The dichotomy between critical thinking and intuition is a false one because it privileges thought over action in all situations. The Prole cannot be both robot and trained professional, whether the context is medical triage or classroom practice, and conversely the Archangel cannot be devoid of bias and uncertainty (Kelly & Elliott-Kelly, 2018). That much is obvious from the decades of claim and counter-claim in EER (see, for example, Muijs et al, 2011). Utilitarianism also ignores emotional motivations such as jealousy and generosity (Harsanyi, 1975 & 1977). It demands only that aggregate benefit be maximized; everything else is disregarded. This is reflected in the development of EER in that peer effects like bullying, friendship and altruistic willingness to help others have been largely ignored, with some exceptions (e.g. Kyriakides et al., 2014). There is no recognition given to the impact that these factors have on outputs or outcomes nor of the impact of the act of measurement on the factors in the first place. As Kelly & Elliott-Kelly (2018) put it: the veins of utilitarianism run through the body of EER both in terms of how data is selected, collected and treated, and in terms of what is ignored.

In summary then, utilitarianism in education implies a willingness to disadvantage some pupils for the greater good. The question for EER in Phase Five is whether it can accept such an approach to schooling; namely, that some pupils are treated unfairly because the greater good is served by ignoring their plight. Few educationalists would accept that one pupil’s deprivation should be weighed against another pupil’s achievement or that fairness can be transferred from one pupil to another like a zero-sum financial transaction. Some commentators like Karl Popper (1945) have suggested a workaround - namely, that instead of ‘the greatest happiness for the greatest number’, we should talk instead about ‘the least amount of suffering for the greatest number’ - but this just rephrases, rather than solves, the problem of accepting a policy of aggregation. EER already has a proud history of carrying out robust empirical research so that educationalists can choose the most effective course of action towards an end (quite properly, in a democracy) determined by policy-makers, but in its next iteration EER needs to articulate the moral impulse behind those policies and decisions (Kelly & Elliott-Kelly, 2018). After all, education is a moral not an economic endeavor, so it needs to identify an immutable ethical basis for equity beyond the current vague notion of having more of it.

**EER and John Rawls’s theory of justice**

John Rawls published *A Theory of Justice* in 1971 as an alternative to utilitarianism. It is a practical attempt to address the tension between freedom and equity in a democratic society, and in that sense it speaks to the underpinning objective of EER. In contrast to utilitarianism, which holds to the single universal principle of maximising aggregated utility, Rawls offers no equivalent universal principle because he regarded ‘the correct principle for anything’ as depending on ‘the nature of that thing’, on the actors and on the context (Rawls, 1971: 29-30). Rawls acknowledges that in a democratic society, while people will have different opinions and competing priorities, there can be only one set of laws, and that this poses two challenges:

* The difficulty of having the state exercise coercive power to force everyone to follow the same set of laws. This is Rawls’s Principle of Legitimacy, the test for which is that the law is enacted and enforced in ways that all stakeholders can endorse and has ‘reciprocity’; that is to say, everyone believes that everyone else will also accept enforcement. Everyone still has their own set of beliefs and values – Rawls called these Comprehensive Doctrines – but they are unwilling to impose their own doctrines on others. Instead, they seek out and agree to mutually acceptable rules as long as everyone else does the same and provided no one group imposes its own Comprehensive Doctrine.

* The difficulty of having people willingly obey the law when that law is devised and implemented by a ruling group whose members probably have different beliefs and values. This is Rawls’s Principle of Stability, the test for which is based on his idea of ‘overlapping consensus’. In this concept everyone endorses the same core set of laws, *but for different reasons*. People support their own ideas of equity and justice consistent with their own Comprehensive Doctrines, but the core set of laws is common to each doctrine and is therefore supported by everyone. Rawls regards ‘overlapping consensus’ as a better basis for social stability than ‘balance of power’, but stability is impossible to achieve when there is insufficient overlap between different sections of society or when there is no convergence on what is meant by equity and fairness.

As far as EER’s education policy context is concerned, it is important that these two challenges identified by Rawls are recognised when find their way into the formal schooling system; for example, by supporting schools that promote illiberal ideologies that seek to impose one group’s Comprehensive Doctrine on everyone else. EER needs to be aware philosophically of the dangers and ensure that its effectiveness metrics – for example, using ethnicity to contextualise schools’ value-added - do not offer perverse incentives towards an undermining divergence and intolerance. Rawls’s theory of justice is predicated on what he called a ‘spirit of public reason’; namely, the belief that people will justify their political decisions to one another in a respectful manner and only by referencing *publicly accepted* (and not *personal*) values. A basic Constitution will act as an overarching guide to and guarantor of these publicly accepted values - the right to vote, the right to own property, and so forth – but of course there will always be a tension as people (respectfully) express their preferences between the aspiration to create a fair society and the rights of the individual. As Kelly and Elliott-Kelly (2018) point out, this is readily applicable to public schooling; for example, the right to maintain good schools for everyone, as part of what Rawls calls the Basic Structure, can sometimes be at loggerheads with the right of individual parents to raise their children and to spend their money as they see fit, even (or especially) if doing so benefits their own children at the expense of the system by going to private schools or paying for extra tutoring. For this reason, Rawls sets out, in a very original way, some fundamental principles for the Basic Structure of social institutions like schools and these are discussed below.

**Rawls’s Veil of Ignorance and the Original Position**

Rawls devised a thought experiment in which the principles for his Basic Structure are chosen in a way that forces everyone to choose only those rules that are fair and justifiable to everyone. He proposed a Veil of Ignorance behind which individuals do not know anything about themselves or about their society, so they do not know which choices will affect them positively and which will affect them negatively. Behind this so-called veil, nothing is known about ethnicity, social class, natural ability, intelligence, age, the structure of society or current affairs. Rawls called this baseline of ignorance the Original Position and he saw it as the best method for reaching a ‘reflective equilibrium’ (Mandle, 2009: 17). Everyone in the Original Position knows that their own interests are at risk from the choices they make, so everyone has the incentive to choose principles that protect *everyone’s* fundamental interests. It is similar to the maximin strategy in game theory where a player chooses the action that produces the *best of the worst* possible outcomes. The Veil of Ignorance deprives people of all knowledge about themselves and about society, but everyone is allowed to know ‘scientific facts’; for example, everyone in the Original Position choosing principles for how society runs schools and educates children is allowed to know the scientific findings of EER and the factors that impact on pupil attainment.

Since the actors do not know their own natural inclinations or circumstances, they do not advocate for any one set of abilities or skills over another and this approach is useful in addressing some issues within EER and its policy context. If parents did not know the intellectual ability of their own children, would they favour the expansion of academically selective schools? If parents were unaware of their own financial situation or social status, would they be in favour of fee-paying schools? If they had no knowledge of whether or not their children had learning difficulties, would they favour or oppose more resources being spent on remedial education? Without knowing whether or not their own children would be affected, what would their preferences be in areas like discipline, streaming by ability and the provision of extra curricular activities (Kelly & Elliott-Kelly, 2018)? These are all questions that the utilitarianism of the earlier phases of EER cannot address beyond demanding that *most* people should benefit in aggregate, but which a Rawlsian approach within an ‘equity phase’ *can* address.

**Rawls’s Principles of Justice**

Rawls suggests that under the conditions of the Original Position, the following two principles will emerge consensually from behind the Veil of Ignorance:

* People will agree to guarantee basic democratic freedom for everyone; in other words, that everyone will have an equal right to the largest set of basic liberties compatible with everyone else having a similar entitlement. This principle – Rawls’s so-called Liberty Principle - will be realised by the *political* institutions of society’s Basic Structure and cannot be traded off against other rights or social goods or against economic prosperity.
* People will agree that in order to allow *any* social or economic inequality, the following two conditions *must both* be satisfied:

(i) Everyone must have a fair chance of getting the best positions in society. Equally gifted people with the same willingness to apply those skills and work diligently should have equal opportunity, regardless of social status.

(ii) Inequality in the distribution of primary goods (defined as those things that everyone would want, whatever else they would want) is only justified if it works for the benefit of everyone and for the most disadvantaged; in other words, for inequality to be accepted, everyone, but especially the disadvantaged, must benefit, though perhaps not to the same extent. This is called the Difference Principle and is based on the premise that the distribution of natural skills and abilities is ‘undeserved’. A pupil does not *deserve* greater advantage simply because he or she was lucky enough to be born with certain academic talents. This is not to say that everyone must get the same share of society’s goods or of nature’s bounty, but it means that the distribution of natural ability should be treated as a common asset that should benefit everyone. Those lucky few who are better endowed by nature can use their innate gifts to make themselves better off, as long as they *also* make the disadvantaged better off (Kelly & Elliott-Kelly, 2018).

According to Rawls, both parts of the second principle will be actualised by the *social and economic* institutions of the Basic Structure, which includes schools (although they were not mentioned specifically by Rawls). It is widely accepted that schools should act to increase social mobility and remediate on behalf of pupils from disadvantaged backgrounds, but the position of schools is more nuanced in Rawls’s theory. Inequality is something that should benefit *everyone*, especially the most disadvantaged, so the challenge for educationalists in a Rawlsian paradigm, ironically, is to examine how *advantaged* pupils benefit from remediating on behalf of disadvantaged ones! After decades of policy striving to correct the social injustices of a world where disadvantaged children are largely left behind, this feels counterintuitive, but Rawls demands that we think about our inherited assumptions and prejudices so that we can better understand how to create a more just society; in this case, to think about how ‘bright’ pupils (say) benefit from having less academically gifted pupils receive additional resources and compensations (like getting extra time in examinations, say).[[6]](#footnote-6) This issue has not been addressed to date in EER because the problem is not recognised in utilitarianism, but it does have a philosophical ‘solution’ in a Rawlsian paradigm. Benefits accrue to ‘advantaged’ pupils (and to advantaged sections of society) from the social coherence generated and secured by the fact that disadvantaged pupils get extra help. Cynics might suggest that disadvantaged pupils only receive enough support to maintain them in their subordinate roles, but a more harmonious, less turbulent society is to *everyone’s* advantage educationally, culturally and economically. The same logic applies, say, to mixed-ability teaching: what bright pupils supposedly lose in not being ‘streamed’ they gain from the harmonious atmosphere of the school’s learning environment, which is why it is important to take account of peer-to-peer effects in measuring effectiveness (Kelly & Elliott-Kelly, 2018).

**A response to critiques of Rawls’s theory**

Some critics regard Rawls’s theory as an attempt to treat effort as morally arbitrary, although effort is the one feature of working-class life that provides self-respect for the disadvantaged. Galston (1991), for example, claims that Rawls’s Difference Principle severs the link between the ‘willingness to produce’ and the ‘right to consume’, replacing claims made on the basis of achievement with those based merely on existence (Mandle, 2009: 31). This is a distortion of Rawls’s theory - the Difference Principle is not there to evaluate *individual* shares, but to evaluate *institutional* and *structural* inequalities – but Kelly and Elliott-Kelly (2018) embrace Galston’s critique and suggest that Rawlsianism can be suitably modified by his criticism before being applied to EER. When Rawls suggests that society’s social and economic institutions, like schools, actualise his second principle, EER might add codicils about the *maturity* and *cultural* *context* of the education system in question, particularly when making international comparisons. For example, social mobility might be a political obsession in the UK, but it might not be a priority in developing countries where citizens might accept greater inequality (say) as long as it created jobs or alleviated famine (say), even if that inequality did not benefit the disadvantaged *most*, as Rawls requires. And in adapting Rawls, EER also needs to discuss how it relates to stability, which is an issue considered by the dynamic model of Creemers and Kyriakides (2008) and in earlier EER research (Bosker & Scheerens, 1994; Sammons, Thomas & Mortimore, 1997; Scheerens & Bosker, 1997; Teddlie & Reynolds, 2000). Rawls suggests that his two principles make societies more inherently stable, but he assumes that the societies in question are democratic and it is not clear how this plays out in undemocratic societies like China, which are included (and often lauded) in OECD international educational league tables. Stable dictatorships may not be more desirable that unstable ones for people whose basic freedoms have been traded-off against transient and ecologically catastrophic prosperity.

Other critiques of Rawlsianism mistake its counter-intuitiveness for weakness – for example, his premise that under certain conditions, inequality is ‘acceptable’ – but this is to lose sight of the fact that opposing philosophies, such as utilitarianism, create inequality under *all* conditions without any preconditions or in the case of communism, strip people of the freedom that Rawls is seeking *a priori* to extract and establish in society as a fundamental human entitlement. The modern utilitarianism of transnational organisations that encourage international competition in education as both participant and referee is a powerful mixture of erudition and hypocrisy designed to obfuscate its own shortcomings, as we will see below with the Restricted Utility Principle in utilitarianism. It survives and has acquired the appeal of a religious sect because it privileges the liberal metropolitan elite who give thanks for their inherited freedoms by whining about them, but have the cultural and economic capital to survive the confusion. With Rawlsianism, on the other hand, what you see is what you get. Rawls simply argues that under his principles people would prefer to maximise the *minimum* amount of benefit that everyone gets, instead of maximising the average amount of primary goods that they receive under utilitarian principles (Freeman, 2003). This is an important point for EER because there is no measure or approach that would be acceptable to international comparison tests which treats success against such a principle; that is to say, that the most successful school or schooling system is the one that maximises minimum (rather than aggregated) pupil achievement. Rawls’s principles secure equal rights for everyone; utilitarianism restricts the basic rights of some for the sake of benefit to the many. Within the sphere of education, the latter permits us to restrict a weak minority or deny them access to schooling – for example, by sending home weak or troublesome pupils during an inspection – if it produces greater utility. This is unacceptable in Rawls’s Original Position. In Rawlsianism, where everyone can see that everyone else has equal basic liberties, pupils are incentivised by the prospect of cooperation based on transparency and mutual respect.

Kelly and Elliott-Kelly (2018) point to another interesting contrast between Rawlsianism and utilitarianism; specifically, between the former’s Difference Principle and the latter’s Restricted Utility Principle, which allows a society to maximise wealth with the only constraint being that the worst-off have a minimum income threshold. In education, most nations guarantee minimum provision for children up to a stated legal school-leaving age and afford extra or compensatory provision for pupils with learning difficulties and for those from poor socio-economic circumstances, while it simultaneously encourages *advantaged* pupils to maximise their attainment at all costs. This is the Restricted Utility Principle in action, but what utilitarianism fails to acknowledge is that those being ‘schooled on support’ will eventually realise that they are being sacrificed to benefit more advantaged pupil and as a consequence will disrupt everyone’s active participation. And in any event, it is not clear that there is any advantage in Restricted Utility; there is no reason why it should deliver greater aggregate utility than Rawls’s Difference Principle. Under Rawlsianism, people can still pursue their own advancement. Rawls’s principles are congruent with self-interest without disadvantaging others, but it is a congruence ‘of the right and the good’ (Mandle, 2009), which requires a sufficient number of people to affirm the same principles of equity in the overlapping consensus. The extent to which this is the case today – i.e. that there is any significant overlapping consensus - in education policy in developed countries is a moot point. The incessant tweaking and upheaval, producing no sustained improvement worth the chaos, is evidence enough of Rawls’s framing of the problem.

It is worth noting that although the Difference Principle depends on the moral claim that it is unfair for people to benefit differentially because of differences between them *that are not their fault*, Rawls does not think that all arbitrary inequalities are unjust. He regarded the ‘natural distribution of talent’ as ‘neither just nor unjust’ (Rawls cited in Mandle, 2009:24), but requires that those who are lucky enough to be born with greater talents – or more accurately, with talents that are in greater demand at the time - are not profiting at the expense ofthose less fortunate, while still being congruent with self-interest. As Rawls (1971: 102) himself says of his own theory, stakeholders ‘agree to share one another’s fate’. This acknowledges the reality of schooling *as a means of social advancement*, but only provided there is sufficient overlapping agreement between different sections of society to affirm the same principles of equity (Kelly & Elliott-Kelly, 2018). This has implications for the adaptation of Rawls to EER and to education generally: how to remediate for those who are born with less academic talent; how to deal with those who can pay for private education; and how to structure learning in schools given natural imbalances in ability (Kelly & Elliott-Kelly, 2018). While Rawls advocates that social institutions like schools should transform the pool of talent into a fair distribution of outcomes, he does not share the same understanding of distributive justice as those who simply believe that fairness requires us to correct all arbitrary inequalities. Rawls’s principles themselves do not require society to even-out handicaps ‘as if it were a horse race’ (Mandle, 2009:25), especially those inequalities that come from natural endowment. Instead, Rawls demands that those who have the same ability and talent, and the same willingness to work diligently to use those talents, should have the same prospect of success. In EER, the narrow definition of ‘success’ that has come from utilitarianism means that within the aggregation of outcomes we have failed to check *who exactly* is achieving *what* in schools, just as we have sometimes failed to check whether or not the school system benefits the most disadvantaged as much as it benefits the well-off (Kelly & Elliott-Kelly, 2018). Part 2 of this chapter addresses those issues and presents new alternative metrics for measuring equity to add to the existing suite of methodologies within EER.

**PART 2: The measurement of equity**

Policy-makers, and transnational organisations like the OECD and the European Union (EU), have encouraged governments to look at educational equity through an economic lens, treating prosperity and consumer choice as desirable outcomes of schooling. Their definitions of educational equity are twofold and can be summarised as: (i) the extent to which pupils can take advantage of education in terms of *opportunities*; and (ii) the extent to which pupils can take advantage of education in terms of *outcomes*. This is a perfectly reasonable view, although there is an intellectual dichotomy contained within it: the former suggests that we look at school effectiveness through the lens of Amartya Sen’s capability theory (Sen, 1982; 1984; 1985a; 1985b); the latter suggests that we look at school effectiveness in terms of attainment and its measurement. While this chapter focuses on measurement, it is worth taking a brief detour to consider the part that Sen’s capability theory might play in the new paradigm and how it dovetails a Rawls’s approach to equity as justice.

In capability theory, equity is about opportunity and taking advantage of it – in fact advantage is a way of viewing relative opportunity - but this is not to be judged solely by pupil attainment. It is possible for a student to have real advantages butnot to make good use of them, and it is possible to have opportunity but not to achieve. Opportunity is not simply whether, say, a pupil can get into a over-subscribed school, but whether the pupil can benefit from the curriculum on offer and the learning atmosphere there (see Kelly 2012a). And acquiring a particular education does not predetermine what a pupil can *do* with it, so as we search for greater equity, some cognisance must be given to what students will actually succeed in *doing* with that greater equity; that is to say, we need to look at pupil ‘functionings’ and personal achievements. Therefore defining equity in education as the extent to which pupils can take advantage of school in terms of *opportunities* captures the freedom of pupils (and parents) to choose functionings *that they value*, and this is in contrast to the classic utilitarian EER paradigm, which defines effectiveness in terms of outcomes and attainment. The utilitarian paradigm, as discuessed in Part One of this chapter, does not distinguish between functioning and capability. For example, entitlement to Free School Meals (FSM) is the most frequently used proxy metric for socio-economic deprivation in school effectivness research, but being entitled to FSM, which is a functioning, is not the same as the capability of being able to eat FSM without feeling stigmatised. And having a more equitable school system might increase social mobility and the ability of pupils to live a better life, but it is more complicated than that: greater equity needs to be accompanied by (among other things) a raising of expectation among students. Otherwise, those from poor backgrounds, who through fate rather than design occupy the margins of society, become reconciled to under-achievement, as do their teachers; they become habituated to the adverse conditions that induce them to accept and endure their lot.

Returning now to the topic of measuring equity. International studies like the OECD’s Programme for International Student Assessment (PISA) attempt to gauge success *at system level* and to facilitate comparisons between nations. To do this, the definition of an equitable system must be narrowed to one in which pupil attainment is independent of those factors that lead to disadvantage, like gender, ethnic minority status, disability, and so forth (see EU, 2006: 2). However, they do not use metrics that allow between-school or within-school comparisons to be made, and ultimately they ignore those aspects of equity - like capability, opportunity and democratic access - that go to the heart of well-being. For example, Shanghai is lauded by OECD for its high attainment in mathematics without acknowledging that China is a totalitarian state on the verge of ecological meltdown whose citizens are subject to continuous political repression. Singapore is also singled out for praise, but Singaporeans live without many of the liberal democratic rights that citizens in Western countries take for granted. The criterion for educational success in OECD terms is clearly driven by *economic* imperatives, which is what China and Singapore have in common, but there should be more to educational equity than this ‘at-least-the-trains-run-on-time’ mentality. Education is more than the training of compliant units of production and there is more to equity than the instrumentalist view that a lack of it inhibits free trade and economic growth. There are *normative* reasons why equity matters, which involve value judgments regarding democratic entitlement and moral purpose, and which go to the heart of the different notions of justice that exist across nations. For Western nations, it *may* be that the price of political freedom is a less efficient schooling system that produces vigilant citizens capable of selecting and de-selecting their own governments. For developing economies, the imperatives may be otherwise. It is important to state this contextual and philosophical limitation clearly here, in advance of describing methods for measuring equity. As this author has noted elsewhere (Kelly 2015), historic attempts to increase equity have drawn heavily on compensation for disadvantaged groups and have been hampered by the perceived need of policy-makers to link greater equity (for pupils) to greater accountability (for teachers). So after decades of focused policy in this area, it is *still* not clear what the outcome targets should be or how attainment should be spread across the range of factors that impact on it. For example, in England, there is not one single reference to ‘equity’ or ‘equitability’ or ‘equality’ in official school improvement reports like those from the London Challenge (DfES 2003, 2006a & 2006b; Ofsted 2010). It is not even clear what policy-makers mean by ‘social justice’, although their desire for it is what drives the equity agenda (Sammons, 2008). How does this relate to utilitarianism? Well, the utilitarian approach to EER happily dodges these issues by concentrating on the measurement of attainment without regard for what it means for the student in terms of leading a ‘better life’ - perhaps this is why the approach is so popular with policy-makers - but neither Sen’s capability approach nor Rawls’s theory of justice can avoid such consideration. Rawls views a just society as a mixed-motive game where individuals are neither totally selfish nor totally unselfish, but where the dominant inclination is to advance one’s own aims through cooperation and agreement rather than through competition and conflict. And capability theory has formal links to Rawlsianism. In fact, Sen’s 2009 book *The Idea of Justice* is a revision of Rawls’s basic ideas, although it criticises the latter for assuming that social justice is a binary, rather than a continuous, variable. According to Sen, and the approach suggested here for a fifth phase of EER, is that social justice and equity are not things that either exist or do not exist. Multiple conflicting but fair principles can co-exist in an equity paradigm, but the main focus for EER should be that aspect of equity that relates to taking advantage of education in terms of outputs and outcomes. So we will now look at six metrics that have been developed over recent decades to do just that.[[7]](#footnote-7)

**1. The Range Ratio and its variations**

Range is the difference between the highest and lowest values of a given variable. It is the simplest measure of dispersion, but it is limited by the fact that it uses only two values from the data set and is greatly affected by extreme outliers. Range Ratio is an improvement on simple range because although it still uses only two data points, at least it ignores outliers. It is calculated by dividing the value at a certain percentile above the median, by the value at a certain percentile below the median.

The Range Ratio is most often used for measuring equity in terms of financial expenditure. A version of it commonly used in the US is the Federal Range Ratio, which divides the difference between spending on the pupil at the 95th percentile and the pupil at the 5th percentile, by the spending on the pupil at the 95th percentile.

 (Spending at 95th – Spending at 5th) **/** Spending at 95th

Another variation is the Inter-Quartile Range Ratio, which is obtained by dividing the spending on the pupil at the 75th percentile by that on the pupil at the 25th percentile.

Spend at 75th **/** Spend at 25th

Both of these measures could be adapted for use with pupil attainment data, but it would require examination *grades* to be converted to *points* or kept as raw percentages.

In the case of the Federal Range Ratio and the Inter-Quartile Range Ratio, and the Palma Index which we will discuss later, the larger the ratio, the *lower* the equity. The lower limit of 1 occurs when the numerator and the denominator are equal. i.e. when there is zero disparity between the cohorts being measured.

**2. The Coefficient of Variation**

The Coefficient of Variation (CoV) is calculated by dividing the standard deviation by the mean.

CoV = σ **/** μ

Unlike the various range ratios described above, the Coefficient of Variation does not depend on just two data points but takes into account *all* areas of a distribution. The higher the CoV, the less equitable the distribution. The lower limit, which represents perfect equity, is zero.

The Coefficient of Variation represents the spread in the data: when the data is not spread out, the peak is high and CoV is small; a distribution that is more dispersed with a lower peak has a higher CoV representing a less equitable distribution. It should be noted that CoV tends to be biased on the low side (i.e. CoV tends to indicate more equity) when *sample* data is used.

**3. The McLoone Index**

Like the Range Ratio, the McLoone Index is most often used for measuring equity in terms of financial expenditure and it is the preferred metric when the lower part of the distribution is of interest (Kelly, 2015). It is calculated using the formula:

Σ(spending < the median) **/** [(Number < the median)×(the median spend)]

The McLoone Index increases as the distribution becomes *more* equitable. Data above the median is not used, but the formula does use a relatively large amount of data and not just two values. The lower limit is zero when the distribution is very *in*equitable and the population below the median receives nothing. The upper limit is 1 when the distribution is perfectly equitable and everyone receives the median amount.

One point to note on the use of the McLoone Index is that the above formula assumes that the disadvantaged group has a *below*-the-median amount of whatever is being measured – usually financial expenditure – so when the disadvantaged group has *above*-the-median amount, it is necessary to invert the McLoone Index ….. or invert the variable (Kelly, 2015); for example, from *pupil-teacher* ratio, say, where lower is better, to *teacher-pupil* ratio where higher is better.

**4. Theil’s T**

Like the various Range Ratios, the Coefficient of Variation and the McLoone Index, Theil’s T has been used traditionally to measure equity in financial expenditure. It is calculated using the equation:

 n

T*indiv* = Σ [(1/n).(v*i*/μ).**ln**(v*i*/μ)]

 *i*=1

where n is the number of individuals in the population, *vi*is the value of the variable (usually financial expenditure) for person *i*, and μ is the population mean (Theil, 1967).

v*i*/μ is the ratio of individual amount to population-average amount, and its natural logarithm determines whether that individual Theil element is positive (when the individual’s amount is greater than the mean), negative (when the individual’s amount is less than the mean) or zero (when the individual’s amount is equal to the mean) in the case of perfect equity (Kelly, 2015). When there is perfect equity and every individual’s amount is equal to the mean, Theil’s T is at its lower limit of zero. When one person has everything and the distribution is totally inequitable, the above formula reduces to

T = {0}+{0}+…+{(1/n).[v/(v/n)].**ln**[v/(v/n)]}

= **ln***n*

It is possible to compare Theils for two or more schools or groups of schools, but even with everything else equal, schools (or groups of schools) with more pupils have higher Theil upper limits, so it is difficult to draw any firm conclusions. Theil’s T is at its best when we are looking at trends over time (Kelly, 2015).

Theil’s T can have both within-school and between-school components. For *n* schools (rather than *n* individuals), T measures equity *between schools* using the equation:

#  *n*

#  Tbet-sch = Σ [p*i*(μ*i*/μ).ln(μ*i*/μ)]

#  *i*=1

where μ*i* is the arithmetic mean of group *i*, p*i* is the fraction of the population in school *i* (the equivalent of 1/n in the ‘individual’ T equation above) and μ is the population mean as before. This is the ‘*between-school’* Theil’s T, which is the T that would be obtained if every pupil in each school had that school’s average share (Kelly, 2015).

The ‘*within-school’* Theil’s T is given by:

#  *n*

#  T*in-sch* = Σ (p*i*μ*i* /μ).T*i*

#  *i*=1

where T*i* is the T of school *i*.

The *overall* T for the whole distribution is given by the between-school and within-school terms added together (and for groups of schools, the ‘between-school’ component will be the lower limit of the overall Theil’s T).

#  *n n*

# T*overall* = Σ [p*i*(μ*i*/μ).ln(μ*i*/μ)] + Σ (p*i*μ*i* /μ).T*i*

#  *i*=1 *i*=1

#  *i*=1

**5. The Attainment Equity Index**

Unlike the previous four measures, which have their provenance in the fair allocation of financial resources, this equity metric was developed specifically for use with pupil attainment data (Kelly, 2012b). It is a Gini-type index based on the assumption that each percentile range of the population of a given school, as measured by prior attainment, achieves the equivalent percentile range of the school’s examination success; that is to say, that a given proportion of a school’s examination grades is attributable to an equal proportion of the pupil population. The evidence from the literature suggests that this is how equity is understood by policy-makers in the UK, the US and Europe (NCLB 2001; Baker & O’Neil 1994; EU 2006).

Figure 2 explains the basis for Gini-type measures. The straight line *y=x* represents ‘perfect’ equity – in other words, the *ideal* distribution of variable *y* over the population *x* - and the curve, called a Lorenz curve, represents the *actual* distribution. A Gini Coefficient is defined as:

A/(A+B)

= 1-2B for normalised axes.

# where *B* is the area under the curve and *A* is the area between the straight line and the curve.

**B**

**A**

**Y**

20

40

60

80

**X**

Cumulative % of population

Cumulative % of variable

Line of ‘perfect’ equality

Lorenz curve

Figure 2. The Lorenz curve and the Gini ‘area’

The Lorenz curve, which defines all Gini-based metrics, represents the proportion of a variable *y* that is cumulatively attributable to the population *x*. If the Lorenz curve is represented by the function *y = L(x),* the Gini can also be given by the formula:

 1

 1 – 2 *∫ L(x) dx*

 0

To integrate the Lorenz function, *L(x)*, in the above formula, the equation must be known, which is seldon if ever the case. So for an Attainment Equity (*Æ*) index trapezoids can be used to approximate Area B (see Figure 3) and use the formula:

 *n*

*Æ* = 1 - ∑ (Xk - X*k-1*)(Y*k* + Y*k-1*)

 *k*=1

where (X*k*,Y*k*) are the points on the Lorenz, with X*0* = Y*0* = 0 and X*n* = Y*n* = 1.

**X**

1

1

Intended

**B**

**A**

**Y**

Cum. % of pupil population

Cumulative % of GCSE A\*-C

Actual

Figure 3. The Lorenz curve approximated

In the example shown on Figure 3, the X*k* points on the horixontal axis are every 20%, so that *Xk* - *Xk-1* = 0.2, but it does not appreciable affect the calculation of the *Æ* Index if the intervals are every 10 percent (say) instead of every 20 per cent, although as Kelly (2015) has noted, most measurements of this kind are lowered by lower ‘granularity’.

# The Attainment Equity Index has certain advantages over the previous metrics: it uses ‘raw’ data; it incorporates all data and not just the extremes; it is based on the well-respected Gini concept; it is easily interpreted and can track changes over time; it is well-suited for use with ratio data like examination results; and as Kelly (2015) has suggested, it can be combined with existing value-added measures to categorise schools in terms of both equity *and* other school effects.

**6. The Palma Index**

Recently, Cobham and Sumner (2013) developed an equity metric for income inequality that could easily be adapted for use with pupil attainment data, and which they claim is superior to Gini-based metrics like Kelly’s Attainment Equity Index. They called it the ‘Palma’ in honour of the Chilean economist Gabriel Palma (2011) who discovered that in nearly all countries in nearly all circumstances, that the 50% of people lying between the 40th and 90th deciles earn approximately half of all national income, and that this proportion is exceptionally stable.

The Palma Index is a type of Range Ratio, similar to the Federal Range Ratio and the Inter-Quartile Range Ratio discussed at the start of this section. Specifically, it is the ratio of income of the top 10% to the bottom 40%. For example, if the richest 10% earn 45% of a country’s income and the poorest 40% earn 20% of the national income, the Palma Index is:

P = 0.45 ÷ 0.2 = 2.25

which means that the richest 10% of the country earn two-and-a-quarter times the income of the poorest 40%.

Like all range ratios, the larger the Palma the greater the *in*equity. The lower limit occurs when the poorest 40% earn virtually all of the ‘other’ 50%; the upper limit is +∞ and occurs when the poorest 40% earn virtually nothing.

Cobham and Sumner (2014) suggest that the Palma Index has advantages over Gini-based indices like the Attainment Equity Index. Gini-based indices do not isolate *where* inequality lies - whether it is in the middle of a distribution or in its tails – and this makes it less than helpful for policy-makers. The Palma, on the other hand, focuses precisely on the tails (as it defines them) and regards inequality as a question of how the ‘other half of the variable’ is proportioned between the richest and the poorest. This makes clear to policy-makers what needs to change; namely, to raise the share earned by the bottom 40% at the expense of the top 10%. Gini-based indices like the Attainment Equity Index are also most sensitive to changes around the *middle* of a distribution and are least sensitive to changes in the tails, but the tails are usually the areas of greatest concern economically and educationally, and this is where the Palma is focussed.

Political policy battles are usually fought along the fault-lines of the ‘haves’ and the ‘have-nots’ and the Palma fits well into this context, but there are several shortcomings with the Palma as far as educational effectiveness research is concerned. The most critical of these is the question of whether or not the Palma’s underpinning assertion holds true that the five deciles from 5 through 9 account for half of all examination success (however that is defined). Given that most assessment in western countries is criterion-referenced rather than norm-referenced, this seems unlikely. Secondly, although the Palma makes clear that what needs to change is the share earned by the bottom 40% at the expense of the top 10%, it is not clear that this can be imposed by policy-makers in schools, since every other government policy seems to incentivise teachers to triage not the bottom 10% but those on the cusp of certain pass grades at whatever decile they occur.

**Technical properties of equity metrics**

# Choice of metric should depend on policy criteria as much as on technical criteria, and on the view taken of the marginal utility value of attainment. At the start of this chapter, utility was defined as the sum of all the benefit (minus the detriment) that results from an action, but it can also refer to the satisfaction obtained from consuming a good or a service. *Marginal* utility, by extension, is the satisfaction a consumer gains from consuming *more* of the good or service. It is not clear whether educational attainment has a positive or negative marginal utility - that is to say, whether the satisfaction gained from greater examination success is decreasing or increasing – but we can say that when attainment has positive marginal utility, the Coefficient of Variation should be used because it has a ‘flat’ response to transfer; and when attainment has negative marginal utility, Theil’s T should be used because transfer among low achievers is more important and Theil’s T is more sensitive at that end of the spectrum. The Attainment Equity Index should be used when one is concerned with *changes* in equity or for middle-ranking schools and schools with comprehensive intakes.

# Kelly (2015) has listed two other important and desirable characteristics of equity metrics. Firstly, they should be scale invariant so that multiplying by a constant or changing units should leave the results unchanged. Fortunately, the Coefficient of Variation, Theil’s T and the Attainment Equity Index can be made scale invariant simply by dividing by the mean. Secondly, equity metrics should be sensitive to changes in the data and they should be transferable; that is to say, equity should be shown to decrease when attainment is transferred from someone with less of it to someone with more of it. The Coefficient of Variation is *equally* sensitive to all transfers, which means it is very *in*sensitive as a metric; whereas Theil’s T is more sensitive at the lower end of attainment, which means that it tends to *underestimate* inequality in underperforming schools because they are more likely to have larger homogeneous populations of ‘have-nots’.

**Conclusions**

‘Effectiveness’ is not a neutral term. Defining it for a particular school will ‘always require choices among competing values’ and an acknowledgement that ‘the criteria will be the subject of political debate’ (Firestone, 1991, p.2). This chapter argues that EER faces challenges regarding its lack of a coherent underpinning philosophy and standing up to those challenges takes the field into a new welcome fifth phase of development. There is little in the early literature to suggest that EER ever felt the need for a formal philosophy as such, so today there is little or no shared understanding within the field of what is meant philosophically or methodologically by ‘fairness’, ‘justice’ and ‘equity’. The two parts of this chapter have addressed these - philosophy and methodology - in turn.

In many respects, EER has been corralled by the defining characteristics of utilitarianism; namely, the aggregation of utility (which has meant discounting pupils whose benefit is below average or accepting that one pupil’s deprivation can be ignored because of another’s achievement) and the primacy of a spurious calculus (which has encouraged the field to measure intangibles in an inappropriate fashion or to ignore factors that are difficult to measure). Part One argued that this utilitarian paradigm is outdated in the era of dynamic models (Phase Four) and effectiveness-for-equity (Phase Five) because it ignores the evolutionary basis of the empathy that people feel for each other. Those who have worked in schools know that pupils often sacrifice self-interest for comradeship, helping across the cognitive, conative and affective domains with academic work, socialisation and acceptance, although a review by Gray (2004) indicated that school effects were typically weaker for affective outcomes than for academic, attendance and behavioural outcomes. Utilitarianism is a credible philosophy with a distinguished provenance, but its shortcomings make it *unsuitable* for EER today because of the moral nature of educational equity. It fails to uphold the intrinsic value of the individual above and beyond the collective, and as such it undermines democracy, social justice and the educative imperative. In its stead, this chapter has argued for Rawls’s theory of justice as the underpinning philosophy of EER in Phase Five and we have seen how Rawls’s Veil of Ignorance approach could help the field theorise about issues like the expansion of academically selective (in the UK, ‘Grammar’) schools (Asthana & Campbell, 2017), the fair allocation of increasingly scarce resources, streaming by ability and the welfare of high-achieving pupils in the state sector. These are issues that utilitarianism has failed to address properly because they are *philosophical* rather than *evidential* issues, and although EER has first-class methodological equipment, it does not yet have the philosophical equipment to tackle them. Emerging from this, Kelly and Elliott-Kelly (2018) have suggested a Rawlsian manifesto for EER to realign it without losing any of its methodological advantages. Such a manifesto would serve to underpin the dynamic approach of Creemers and Kyriakides (2008) by adopting the following five principles for Phase Five:

1. *The Benefit Principle:* Educational effectiveness, which is multi-level in nature and dynamic in how it changes over time, is that which increases educative benefit for all pupils, but increases the benefit for disadvantaged pupils more. Inequality is permitted and accepted only if it benefits everyone and especially the most disadvantaged.
2. *The Redress Principle:* Schools – even effectives ones - cannot correct for the fact that some children get a smaller share of society’s goods or of nature’s bounty, but schools can, and should, make a significant contribution towards redressing undeserved imbalances; that is to say, effective schools have an equity imperative.
3. *The Immutability Principle and the Within-School Imperative:* An effective school does not trade-off one pupil’s failure for another’s achievement. The metrics to measure effectiveness may include aggregate measures at the level of the school, but they should look primarily at specific educative benefit at the level of the pupil and the classroom. The primary objective is to gauge ‘*who’* is achieving ‘*what’*. The metrics to gauge *systemic* effectiveness underpin a secondary objective in looking at aggregated benefit and between-school performance, and this should take account of grouping at the meso-level between the institution and the system, such as with ‘chains’ of schools and academy ‘trusts’.
4. *The Veil of Ignorance Principle applied to Policy:* Controversial national policies should be considered using Rawls’s Veil of Ignorance approach, with new experimental methodologies as required, to enable stakeholders to give preferences free from bias and self-interest.
5. *The No-Harm Principle.* All methodologies, including those that enable benevolent and malevolent peer effects to be included, must pass a ‘no-harm test’ to ensure that they do not adversely affect the overlapping consensus.

Of course, declaring one school or system more equitable than another school or system depends not just on measurement, but on one’s definition of equity. Sammons (2007:20) has suggested that ‘judgements about school effectiveness need to address three key questions essential to the consideration’ and promotion of social justice: the ‘what’ of effectiveness (which outcomes?); the ‘who’ of effectiveness effective (which student groups?); the ‘when’ of effectiveness (over what time period?). These questions provide a sound basis for monitoring both an education system and an individual school’s success in promoting equity and equal opportunities for all its students. They can also provide a clear focus for school development and improvement, planning and evaluation. More genereally, policy-makers in education see equity – more specifically, *attainment* equity – as a mechanism for creating greater social mobility and by inerence, a more just society. Rawls’s contribution in this respect fits well within the EER paradigm, although his theory of justice is not without its detractors. The assumed causal link between equity and opportunity is problematic too. Equality of opportunity does not necessarily result in equality of outcomes - in some contexts, equality of opportunity may actually accentuate *in*equality - and in any case, having more opportunity does not always result in greater achievement. Tristram Hunt (2015), the well-respected former Labour Party shadow Secretary of State for Education in the UK, warned that ‘if Labour were ever to abandon equality, there would be very little left to distinguish it from the Conservative Party’. This was a view originally put forward by Hugh Gaitskell, the former (1955-1963) Labour Party leader, but the truth of the matter, as Hunt acknowledges, is that we have already arrived at a policy juncture where the Right is thought not to care about equity and the Left doesn’t talk about it. The UK is the most unequal OECD country in Europe: a country where, despite decades of what Hunt calls ‘a dizzying array of micro-targeted’ education policies, there is a generation growing up with less opportunity and lower achievement than their parents.

“*The barriers that bright, wonderful, earnest five year-olds face in modern Britain are wretched to behold. … That’s the truth about inequality the statistics don’t always show. The political philosopher Roberto Unger has described the modern centre-Left as ‘content to appear … as humanisers of the inevitable’. Sadly, it is not an altogether unfair description.*” (Hunt, 2015)

Education policies in relation to equity, especially those of western democracies, have failed and if ‘the truth about inequality’ doesn’t always show up in the statistics as Tristram Hunt says, that is a measurement problem. Whatever our philosophical position on equity and justice, and whatever its cause, measuring and monitoring equity can help gauge the effectiveness of policies aimed at increasing it (Kelly, 2015). This is the motivation behind this chapter: we need to know which pupils are achieving the appropriate grades, why there are gaps in achievement and whether, for example, equity might more easily be increased in norm-referenced examination systems. It is an urgent problem: the trend is towards ever-greater inequality. As Hunt (2015) says, it is ‘staggering how much schools have to swim against the current when it comes to helping poorer children fulfill their potential’.

**References**

Alican, N. (1994). *Mill’s Principle of Utility: a defense of John Stuart Mill’s notorious proof*. Atlanta GA. & Amsterdam: Rodopi.

Asthana, A. & Campbell, D. (2017, March 6) Theresa May paves way for new generation of grammar schools. *The Guardian.* Retrieved from https://www.theguardian.com /uk-news/2017/mar/06/theresa-may-paves-way-for-new-generation-of-grammar-schools [Accessed 30 March 2017]

Baker, E. and O’Neil, H. (1994) Performance Assessment and Equity: a view from the USA, *Assessment in Education: Principles, Policy & Practice,* 1(1), 11-26.

Bentham, J. (1780). *An Introduction to the Principles of Morals and Legislation*. (2007 edition, which is an unabridged republication of the Clarendon Press 1907 edition.) Mineola NY: Dover Publications.

Bosker, R. & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. *International Journal of Educational Research,* 21, 159-80.

Cobham, A. and Sumner, A. (2013) *Is it all about the tails? The Palma measure of income inequality,* CGD Working Paper 343 (Washington DC, Center for Global Development).

Cobham, A. and Sumner, A. (2014) Is inequality all about the tails? The Palma measure of income inequality, *Significance*, February, 11(1), 10-13.

Coleman, J.S., Campbell, E.Q., Hobson, C.F., McPartland, J., Mood, A.M., Weinfeld, F.D., &

York, R.L. (1966). *Equality of Educational Opportunity*. Washington, DC: US Government

Printing Office.

Creemers, B. (1994). The Effective Classroom. London: Cassell.

Creemers, B. & Kyriakides, L. (n.d.) Using educational effectiveness research to design theory-driven evaluation aiming to improve quality of education. Available online at: http://www.ivalua.cat/documents/1/19\_10\_2011\_15\_49\_19\_Ponencia\_Taula1\_Creemers\_Kiriakidis.pdf

Creemers, B. & Kyriakides, L. (2006) Critical analysis of the current approaches to modelling educational effectiveness: the importance of establishing a dynamic model, *School Effectiveness and School Improvement*, 17(3), 347-66.

Creemers, B. & Kyriakides, L. (2008). *The Dynamics of Educational effectiveness: a contribution to policy, practice and theory in contemporary schools*. Abingdon: Routledge.

De Jong, R., Westerhof, K.J., & Kruiter, J.H. (2004). Empirical evidence of a comprehensive

model of school effectiveness: A multilevel study in mathematics in the 1st year of junior

general education in the Netherlands. *School Effectiveness and School Improvement*, 15(1),

3–31.

DfES (2003) *The London Challenge: Transforming London Secondary Schools* (Nottingham, DfES publications).

DfES (2006a) *London Challenge: Families of schools, May 2006, secondary schools* (London, Department for Education and Skills).

DfES (2006b) *London Challenge: Families of schools, May 2006, primary schools* [in 4 volumes] (London, Department for Education and Skills).

Edmonds, R. (1979) Effective schools for the urban poor, *Educational Leadership*, 37(1), 15-24.

EU (Commission of the European Communities) (2006) *Communication from the Commission to the Council and to the European Parliament: efficiency and equity in European education and training systems*. (Brussels, 8.9.2006, COM2006 481 final)

Firestone, W. (1991). Introduction: Chapter 1 in: J. R. Bliss, W. Firestone & C.

E. Richards (Eds). *Rethinking Effective Schools: Research and Practice*

(Englewood Cliffs, New Jersey: Prentice Hall).

Freeman, S. (Ed.) (2003). *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press.

Galston, W. (1991). *Liberal Purposes: goods, virtues and diversity in the liberal state*. Cambridge: Cambridge University Press.

Gray, J. (2004) School effectiveness and the ‘other’ outcomes of secondary schooling: a reassessment of three decades of British research, *Improving Schools*, 7(2): 185-198.

Hare, R.M. (1973) The Presidential Address: principles, *Proceedings of the Aristotelian Society, New Series*. 73, 1-18.

Hare, R.M. (1981). *Moral Thinking: Its Levels, Method and Point.* New York: Oxford University Press.

Harsanyi, J. (1975) Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory A Theory of Justice by John Rawls, *The American Political Science Review*, 69(2), 594-606.

Harsanyi, J. (1977) Morality and the Theory of Rational Behavior, *Social Research*, 44(3), 623-56.

Hunt, T. (2015) *Labour’s Moral Mission: Inequality and Social Mobility in Modern Britain.* Speech to the Fabian Society,  7 December 2015. Private communication with the author.

Hutcheson, F. (1726). *An Inquiry into the Original of Our Ideas of Beauty and Virtue*. (Liberty Fund 2004 edition, edited by W. Leidhold.) Indianapolis: Liberty Fund. Retrieved from http://oll.libertyfund.org/titles/hutcheson-an-inquiry-into-the-original-of-our-ideas-of-beauty-and-virtue-1726-2004 [Accessed 30 March 2017]

Kelly, A. (2012a) Sen and the Art of Educational Maintenance: evidencing a capability, as opposed to an effectiveness, approach to schooling, *Cambridge Journal of Education*, 43(3), 283-96.

Kelly, A. (2012b) Measuring ‘equity’ and ‘equitability’ in school effectiveness research, *British Educational Research Journal*, 38(6), 977-1002.

Kelly, A. (2015) Measuring Equity in Educational Effectiveness Research: the properties and possibilities of quantitative indicators, *International Journal of Research & Method in Education*. 38(2), 115-136

Kelly, A. & Elliott-Kelly, C. 2018. Towards a Philosophy of Equity in Educational Effectiveness Research: moving from utilitarianism to a Rawlsian paradigm, *School Effectiveness and School Improvement*.

Kragh, H. (1990) *Dirac: A Scientific Biography* (Cambridge: Cambridge University Press).

Kyriakides, L., Campbell, R.J., & Gagatsis, A. (2000). The Significance of the Classroom Effect

in Primary Schools: An Application of Creemers’ Comprehensive Model of Educational

Effectiveness. *School Effectiveness and School Improvement*, 11(4), 501–529.

Kyriakides, L., Creemers, B., Papastylianou, D. & Papadatou-Pastou, M. (2014) Improving the School Learning Environment to Reduce Bullying: An Experimental Study, *Scandinavian Journal of Educational Research*, 58(4): 453-478.

Mandle, J. (2009) *Rawls's A Theory of Justice: An Introduction*. Cambridge: Cambridge University Press.

Mill, J.S. (1863) *Utilitarianism.* Retrieved from https://www.utilitarianism.com/mill2.htm

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988) *School Matters: The Junior Years* (Wells: Open Books).

Mortimore, P. & Whitty, G. (1997) *Can School Improvement Overcome the Effects of*

*Disadvantage?* (London: Institute of Education University of London).

Muijs, D., Kelly, A., Sammons, P.**,** Reynolds, D. &[Chapman, C.](http://eprints.gla.ac.uk/view/author/29229.html)  (2011) The value of educational effectiveness research: a response to recent criticism, [*Research Intelligence*](http://eprints.gla.ac.uk/view/journal_volume/Research_Intelligence.html), 114: 24-25.

NCLB (No Child Left Behind) Act (2001) *Public Law print of PL 107-110*. Available online at: http://www2.ed.gov/policy/elsec/leg/esea02/index.html

Accessed 10 October 2014

Nuttall, D. (1990) *Differences in Examination Performance RS* 1277/90 (London:

Research and Statistics Branch, ILEA).

Ofsted (2010) *London Challenge - Reference no. 100192* (London, Ofsted).

Opdenakker, M.-C. & Van Damme, J. (2000) Effects of schools, teaching staff and classes on achievement and well-being in secondary education: similarities and differences between school outcomes, *School Effectiveness and School Improvement*, 11(2): 165–96.

Palma, J. G. (2011) Homogeneous middles vs. heterogeneous tails, and the end of the ‘inverted-U’: it’s all about the share of the rich, *Development and Change*, 42(1), 87–153.

Popper, K. (1945). *The Open Society and its Enemies.* London: Routledge. In two volumes. [Reissued in one volume as Popper, K., Ryan, A. & Gombrich, E. (2013). *The Open Society and Its Enemies*. Princeton, NJ: Princeton University Press.]

Rawls, J. (1971). *A Theory of Justice* (1999 edition). Cambridge, MASS: Harvard University Press.

Reynolds, D**.,** Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C. & Stringfield, S. (2014) Educational effectiveness research (EER): a state-of-the-art review, *School Effectiveness and School Improvement*, 25:2, 197-230.

Rutter, M., Maughan, B., Mortimore, P. and Ouston, J. (1979) *Fifteen thousand hours: Secondary schools and their effects on children*. London: Open Books. [Reprinted 1995. London: Paul Chapman].

Sammons, P. (2007). *School effectiveness and equity: making connections*. Reading: CfBT.

Sammons, P. (2008) Zero tolerance of failure and New Labour approaches to school improvement in England, *Oxford Review of Education*, 34(6): 651-664.

Sammons, P., Davis, S., & Gray, J. (2016) Methodological and Scientific Properties of School Effectiveness Research: Exploring the underpinnings, evolution and future directions of the field, in:C. Chapman, D. Muijs, D. Reynolds, P. Sammons & C. Teddlie (Eds) *The Routledge International Handbook of Educational Effectiveness and Improvement: Research, policy, and practice* (London:Routledge International Handbooks of Education) pp. 25-76.

Sammons, P., Hall, J., Sylva, K., Melhuish, E., Siraj-Blatchford, I., & Taggart, B. (2013) Protecting the development of 5–11-year-olds from the impacts of early disadvantage: the role of primary school academic effectiveness, *School Effectiveness and School Improvement*, 24(2): 251-268.

Sammons, P., Thomas, S. & Mortimore, P. (1997). *Forging links: effective schools and effective departments*. London: Paul Chapman.

Scheerens, J. (1992). Effective Schooling: Research, Theory and Practice. London: Cassell.

Scheerens, J. & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon Press.

Sen, A. (1982) Rights and agency, *Philosophy and Public Affairs*,11, 3-39.

Sen, A. (1984) *Resources, values and development* (Oxford, Basil Blackwell).

Sen, A. (1985a) *Commodities and Capabilities* (Amsterdam, North-Holland).

Sen, A. (1985b) Well-being, Agency and Freedom: the Dewey Lectures, *Journal of Philosophy*, 82(4), 169-221.

# Sen, A. (2009) *The Idea of Justice* (Cambridge, Mass., Harvard University Press & Allen Lane).

Sidgwick, H. (1874). *Methods of Ethics* (7th/1981 edition). New York: Hackett Publishing.

Stringfield, S. & Slavin, B. (1992) A hierarchical longitudinal model for elementary school

Effects, in: B. Creemers & G. Reezigt (Eds) *Evaluation of Educational Effectiveness*,

pp. 35–69. Groningen.

Teddlie, C. & Reynolds, D. (2000). *The international handbook of school effectiveness research*. Lewes: Falmer Press.

Theil, H. (1967) Economics and Information Theory (Chicago, Rand McNally).

1. We distinguish between an ‘outcome’ and an ‘output’ although the two are conflated in the literature. An outcome (e.g. a satisfying job or a happy life) is the benefit / consequence of an output (e.g. good examination grades or a university degree). [↑](#footnote-ref-1)
2. For a detailed discussion of the origin and development of the field, see Reynolds et al. (2014) and Sammons, Davis & Gray (2016). [↑](#footnote-ref-2)
3. Philosophical elements of this chapter have appeared previously in Kelly and Elliott-Kelly (2018) and methodological elements in Kelly (2015). [↑](#footnote-ref-3)
4. Interestingly, the seventh edition, Sidgwick, H. (1981) *Methods of Ethics* (New York: Hackett Publishing), has a preface written by John Rawls. [↑](#footnote-ref-4)
5. In his book *Theory of Legislation*, Bentham distinguishes between ‘evils of the first and second orders’ First-order evils have immediate consequences; second-order evils occur when consequences spread through society causing disruption, and it is the latter that ‘makes punishment necessary’. [↑](#footnote-ref-5)
6. It might be useful to distinguish Rawlsianism from traditional egalitarianism at this point. The latter is known for its negative attitude to regulation and its positive attitude towards collective decision-making, so an egalitarian society is motivated by cooperation and peer pressure rather than by competition and regulation. However, modern egalitarianism rejects this, holding that if everyone had the same opportunity cost, there would be no relative advancement and no one would derive any benefit from dealing with others in society. In the egalitarian view, the benefits that people get from dealing with each other arise *because* they are unequal, whether that inequality is from natural or from nurture. [↑](#footnote-ref-6)
7. A more detailed mathematical treatment, with worked examples, can be found in Kelly, 2015. [↑](#footnote-ref-7)