EQUIPOISE AND THE EXPERT COMMUNITY IN RANDOMISED CONTROLLED TRIALS: RESEARCH ETHICS IN 'CLOSING THE GAP'

Nigel Fancourt: University of Oxford

ABSTRACT

This article considers various aspects of research ethics in randomised controlled trials in education. They arose out of one large educational research project: 'Closing the Gap: Test and Learn' which involved seven RCTs with randomisation at school level and 4 replications with randomisation within individual schools – with over seven hundred participating schools. First the article explores how research ethics are addressed in recent policy documents supporting the use of randomised controlled trials in schools. Then there is an account of the ethical issues in the project, and their pragmatic solutions. Five stages are identified: choice of interventions; the use of control groups; data collection; use of interim data; subsequent research. The problems posed by two of these - choice of interventions and sharing of interim data - are then reviewed in detail in the light of the notion of equipoise, and particularly in identifying the appropriate expert community given the range of organisations involved, including schools themselves. The need for more attention to equipoise in educational research ethics, both in terms of governance and deliberation, are discussed, as well as some of its limitations for school-led research.

INTRODUCTION

The importance of addressing the ethical dimensions of any research involving people hardly needs to be highlighted, and ethical conduct in educational research has long been seen as an essential element of current research practice (e.g. Burgess 1989; MacNamee and Bridges 2002). Researchers have explained and reflected upon their ethical experiences and deliberations (Midgley et al. 2013), tacking issues such as: consent (e.g. Robinson-Pant and Singhal 2013; Shamim and Qureshi 2013); anonymity (e.g. Walford 2005); children and young people (e.g. Parsons et al. 2015); the use of images (Kaplan et al. 2011). This collective process of reflection has informed and been informed by the development of codes of practice, notably BERA's guidelines (BERA 2011). In university-based research, there is usually a process of formal scrutiny across many disciplines, though some would argue that ethical reasoning has thereby become over-bureaucratised and researchers' deliberative judgements have been compromised or hidden, both in the social sciences (e.g. Haggerty 2004; Israel and Hay 2006; Hammersley 2010), and indeed in medicine (Chalmers 2011). The ethical issues surrounding research by teachers has also come under scrutiny (e.g. Zeni 2001; Brindley and Bowker 2013; Mockler 2014), as have the ethical issues of collaboration between universities and schools (e.g. Locke et al. 2013).

There have therefore been recent pleas for more openness about ethical deliberation, and thus Parsons et al. (2015) urged that:

> There should be greater opportunities for researchers to share their practices…so that creativity becomes a more transparent and, therefore, accepted part of the ethics landscape (Parsons et al. 2015, p. 725)

With this in mind, this article presents some ethical challenges in one recent major quantitative study. However, it is less a sharing of what the researchers practiced and more a reflection on how it might have been improved, notably by the use of the bioethical notion of equipoise, and thus it seeks to be retrospectively creative. As the use of RCTs, is growing (Connolly 2015), these issues are increasingly in the spotlight, and for instance Fives et al. (2015) recently reviewed some of the ethical issues arising from one educational RCT. Like Fives et al., this paper will address some of the internal tensions in ethical deliberation; however unlike their article, it will not address the wider issue of whether RCTs are ethically permissible at all – in response to, for instance Morrison (2001) or Hammersley (2008).

The 'Closing the Gap: Test and Learn' project was an unprecedentedly large scale randomised controlled trial of seven different interventions, involving five research organisations and seven hundred schools across nearly two hundred alliances (Churches 2016b; see also Churches 2016a; and National College for Teaching and Leadership 2016a, 2016b, 2016c, 2016d, 2016e, 2016f, 2016g, 2016h). The organisational complexity and choice of methodology raised a range of epistemological and pragmatic questions, as discussed by Thompson and Menter (XXXX). It also presented several unusual issues for research ethics, some of which were identified and addressed in advance, whereas others only emerged later in the project. The research report only gave a very brief account of these issues (Churches 2016b, p.38-39), so this article offers a retrospective

review of these issues and how they were addressed, as a reflection, drawing on some internal documents as well as qualitative interviews with various stakeholders, as described in Childs et al. (XXXX), and Thompson and Menter (XXXX). The first section is chronological, identifying and commenting on the project's approach to ethical issues over the course of the research, followed by a review. It is certainly not intended to suggest that the best choices were made, but rather to illuminate the decision-making process, to inform future research design and ethical deliberation in randomised controlled trials in education.

THE ETHICS OF RANDOMISED CONTROLLED TRIALS IN RECENT EDUCATION POLICY

The initial impetus for the CtG project was a desire within the Coalition government (2010-2015) to raise the attainment of pupils who faced educational disadvantages, and this has long been a concern of various political shades. However, it also became a major part of the Coalition's drive to develop the use of randomised controlled trials ('RCTs') both in public policy generally (Haynes et al. 2012) and specifically in education (Goldacre 2013), modelled on their use in medicine and the health sciences. There was however some reaction to the analogy with evidence-based practice in medicine, as educational researchers argued that this methodological approach was neither new nor straightforward (e.g. James 2013; Taber 2013). Further, the policy aimed to enable schools to make research-informed decisions, so that time and money was not wasted on irrelevant or ineffective strategies, on a 'what works' rationale. Put another way, the project was not simply an abstract investigation of interesting research topics; it was both grounded in political and professional concerns about underachievement and educational equity, and also a vehicle for developing research capacity in schools.

A striking feature of the two policy documents was the framing of ethical concerns. In their argument for the use of RCTs in public policy, Haynes et al. (2012) hoped to dispel what they perceived to be the main criticisms of the use of RCTs, in a section entitled 'The case for RCTs – debunking some myths', covering issues of ethics as well as complexity, cost, and need. The term 'myth' is a surprisingly pejorative label for genuine methodological concerns. They focused on the ethics of control groups:

> Sometimes people object to RCTs in public policy on the grounds that it is unethical to withhold a new intervention from people who could benefit from it. This is particularly the case where additional money is being spent on programmes which might improve the health, wealth, or educational attainment of one group. (Haynes et al. 2012, p. 16)

Although none were specifically cited educational researchers do indeed have these concerns, for example Gorard (2001). In reply, Haynes et al. (2012) argued from an analogy with medical research, holding that although it would be wrong to withhold an intervention if its benefits were known, nevertheless 'we need to be clear about the limits of our knowledge and that we will not be certain of the effectiveness of an intervention until it is tested robustly' (p. 17). They then give examples of professional practices which had been considered appropriate until they were shown to be ineffective or even detrimental when tested in RCTs. One of the co-authors, Ben Goldacre, went on to address the educational issues specifically (Goldacre 2013), and unsurprisingly adopted the same language – 'myths about randomised trials' (p. 10) - addressing the related concern that 'people sometimes worry that it is unethical to randomly assign children to one educational intervention or another' (p. 11). He then narrowed this to a particular assumption: 'Often this is

driven by an implicit belief that a new or expensive intervention is always necessarily better' (p. 11), and put forward a justification for RCTs on the basis of the need to tackle the claims of charismatic 'experts' with robust evidence. However, they may have genuine concerns that the intervention is potentially harmful to pupils' learning.

Neither Haynes et al. (2012) nor Goldacre (2013) intended to provide a comprehensive guide to the ethical issues, but their approaches to ethical concerns in these papers raise some questions. This is largely because they simply counter an ethical objection to RCTs with an alternative claim, but do not provide any indication as to how one should balance these two claims; e.g. the potential benefits versus uncertainty of effectiveness. Generally, their enthusiasm for RCTs tended to skip over full consideration of these issues, but an ethical justification for conducting a RCT is not the same as ethically conducting a RCT.

THE ETHICAL ISSUES IN 'CLOSING THE GAP'

The CtG project can be reviewed in the light of the preceding discussion, in order to consider how ethical issues were explicitly and implicitly considered, framed and played out over the course of the research. Much discussion of practice-based educational research has pointed out the ethical complexities of how school work with an outside agency, typically a university (Locke et al. 2013). A significant contextual factor in exploring these issues was the number of different organisations involved, including the government agencies (Department for Education and the National College for Teaching and Leadership), the two universities (Durham and Oxford), and

two independent educational organisations (CfBT and CUREE). Further, there were also hundred and eighty teaching school alliances, and nearly seven hundred schools involved, who were also gathering and using some of the data from their own pupils, and indeed one of the twin aims of the project was to promote research capacity for RCTs in schools. Clearly, there were a range different professional responsibilities and practices at work, which had to be reconciled.

Overall, there were five different stages which raised ethical issues: choice of interventions; the use of control groups; data collection; use of interim data; subsequent research. These are discussed chronologically, to highlight how different issues arose over the course of the project. First, the initial planning stage involved choosing the different interventions to be in the project. This review was carried out by NCTL, Durham and CUREE, and included gathering a focus group of teachers for consultation, over seventeen different meetings. The aim of this process was, according to the final report, to ensure that the interventions were a selection of those strategies 'which the current evidence supported as being effective in closing the attainment gap for lower-performing pupils, with a view to evaluating them using large-scale RCTs' (Churches 2016b, p.19); however the wording of the first clause is imprecise, in that, as a reviewer commented 'we would not have proposed trials for things which were know to be effective for a specific context', and could be interpreted as meaning what current *practice* supported. The selection included different types of strategy, from classroom-based activities, such as Numicon, to whole-school initiatives such as Achievement for All; however they satisfied the two requirements that whilst they were not definitively advantageous, they were also considered by the selection group to be unlikely to be detrimental in comparison to current practice.

The use of control groups was the focus of internal discussion, particularly as schools were thought to be likely to object to being involved on this basis. A member of the research team had a perception that there was 'a general tendency...to dismiss RCTs as...unethical', echoing the points raised in the policy documents (Haynes et al. 2012; Goldacre 2013). This was addressed in the early training rounds by drawing on a medical analogy, and particularly highlighting that the control group schools were not to be conceived as 'doing nothing', instead they were entitled and indeed encouraged to do as much as they could for their pupils, acting on their own professional judgement: the interventions needed to be better than whatever schools could provide themselves. In the event, a member of the research team felt that, 'we were surprised at the relatively low resistance to the concept of a RCT amongst the participants...by and large schools' concerns about randomised controlled trials have been almost entirely pragmatic ones, not the ethical ones'.

Simultaneously, the process for giving consent needed attention. The most rigorous position would have been for explicit parental and individual consent for every pupil (e.g. Homan 2002). This would have been complex given the vast number of schools and pupils involved: the logistics of withdrawing different pupils from the research would be convoluted for classroom interventions such as First Class@Number, and harder again for whole school projects such as Achievement for All. On balance, headteachers' consent was considered appropriate, for several reasons. First, under current legislation teachers have a legal responsibility for pupils' welfare (UK Government 1996), and are therefore deemed to be able to act in the interests of the child. Furthermore, they could also choose to adopt the interventions without parental consent; for example, a school would not need parental or pupils' permission to start using Numicon. Third,

the main sponsor, the Department for Education (through NCTL), already had the right to collect attainment data on all pupils; neither pupils nor parents can withdraw from national analyses of school data, such as GCSE results. While law and ethics are not the same, on balance it was considered appropriate to deem headteachers as the appropriate givers of consent on behalf of the pupils.

A further concern was that the project aimed at addressing the needs of educationally disadvantaged pupils, who could be seen as particularly vulnerable. However, on balance, they were not deemed to require special consent as: they were in mainstream schooling, were not extremely vulnerable in the sense of being physically or emotional at risk, the individual data on each of them was minimal, and their data was often nested within whole class or school data. Further, they were not singled out at the data collection stage, for instance by interviewing them apart from their classmates.

There is no evidence that schools or alliances withdrew on ethical grounds, nor that any parents objected to their child's involvement, though this point was not been fully researched. Data was stored securely by the research team at the NCTL and CfBT in accordance with the Data Protection Act. Durham University carried out some analysis of anonymised secondary data. At the analysis stage, no pupil, school or alliance has been identified; however, as schools and alliances became involved in further projects, such as 'early adopters', their involvement in the project became more visible, through ad hoc self-identification (Churches and McAleavy 2015; NCTL 2016c, 2016d, 2016e, 2016f, 2016g).

A new ethical issue arose at the end of the first year, once the interim data had been analysed. This data showed that the experimental groups for some for the interventions were not as successful as the control groups, particularly for the under attaining pupils - RTI, First Class @Number, and Inference Training (Churches 2016b, p. 47); such pupils in these trials made less progress than their peers in the control groups. The researchers had to decide whether to continue with these interventions at all, and/or whether to disclose the interim analysis to the schools. It might have been possible to halt the RCTs, but not disclose the data, though it would have implied some concerns, especially if some interventions were halted and not others.

An alternative was to continue with the RCTs and to publish the interim data, in order to allow the schools to decide for themselves if they wished to be involved. This could be seen as allowing the schools leaders, as educational experts, to reconsider their consent on behalf of the pupils. This new information could inform their consent, and the decision to treat the headteachers as suitable givers of consent becomes more sharply focused. There was a complex interplay between consent and knowledge. One could argue that it was for the headteachers, as educational experts, to decide whether they still considered there to be uncertainty, given that they were deciding on behalf of the pupils. If they were deemed to be able to act in the child's best interests by giving consent, they arguably they could have been given all the available evidence, if that might affect their decision-making.

The other concern was in terms of ethical obligations to the providers of the interventions and disclosure of the data. This type of relationship is not specifically identified in BERA's guidelines (2011), but is broadly covered by the discussion of the 'community of educational researchers' (p. 9), and includes not presenting a partial or distorted picture, for instance by releasing data or analysis whether positive or negative, before the research was completed. In this case, the research had been planned as a two-year process, and one of the interventions (Achievement for All) was intended to take two years to effect, so it would have been premature to halt it beforehand. However, if this intervention were to continue it could then be considered unfair to halt other interventions which performed similarly, but were only planned to take a year, since they too might improve in the second year.

In the event, it was decided to halt one of the interventions, Inference Training. This showed the largest negative effect, against a control group which made significant progress (Churches 2016b, p. 33). This arguably may have been due to problems of instrument validity, in that the literacy tests measured pupils' comprehension, not ability to draw inferences, and while pupils' ability to infer meaning might have a consequential effect on their comprehension, that would be indirect and might require a longer intervention. This therefore was a new ethical decision within the project, involving balancing the new emerging findings about the effectiveness of the intervention against any ostensible benefits. Schools were not informed of any detailed results, on the basis that the research was incomplete, and that some of the data was missing before the start of the second academic year. They were given a brief summary of the results of the trials. This decision was to some extent justified by the results after Year 2 in that some interventions

which were relatively unsuccessful in Year 1 were successful in Year 2, for instance X and Y, although other interventions remained unsuccessful throughout (Achievement for All).

EQUIPOISE, THE EXPERT COMMUNITY AND INTERIM DATA

The preceding account of the ethical issues raised by the project raises some problems common to all educational research, such as consent or anonymity, but also two problems which are peculiar to RCTs: choice of interventions and the use of interim data. At their heart is a tension between, on the one hand, an entirely ethical question about beneficence and non-maleficence - to do good and not to do harm – and an epistemological question about the certainty of knowledge. This can be framed in classical utilitarian terms, as a process of reconciling remoteness, certainty and extent (Bentham 1996). When selecting the interventions, it was not certain whether any of the interventions were effective, but it was also considered by the reviewing panels of teachers that they would not be harmful, and indeed might be beneficial to their pupils.

In retrospect, it would have been valuable to have looked more closely at the considerable discussion in medical research about the ethics of RCTs. Medical researchers have to ask similar questions when deciding whether to test out a new drug or procedure, and also when they are confronted by interim data halfway through a clinical trial. In these situations they often make use of the principle of *equipoise*. This was initially defined as 'the point where a rational, informed person has no preference between two (or more) available treatments' (Chard and Lilford 1998, p. 891; see Freid 1974; Lilford and Jackson 1995). The principle means that it is

inappropriate to launch into RCT of a new treatment solely on the basis that it would be interesting scientifically; some deliberation of the ethics of the treatment is needed. Moreover, it is inappropriate on the basis of weighing up the benefits and risks of the new treatment alone. Some comparison of the risks and benefits in relation to the other available treatments on offer, especially existing practice, is required.

This raises a question as to who the appropriate 'person' is to make this judgement- the clinician, researcher or patient - and whether this is to be viewed as an individual lack of preference or a collectively, e.g. by the doctor alone, or across current practice. A more nuanced formulation emerged in Freedman's definition conceptualisation of *clinical* equipoise, defined as when "there is genuine uncertainty within the expert medical community — not necessarily on the part of the individual investigator — about the preferred treatment" (Freeman 1987, p. 141). Thus an individual's doubts in the face of current good practice or research evidence are not enough to warrant a randomised controlled trial, but researchers have to consider the views of the professional community collectively. This means that it is unethical to try out a RCT in the face of widespread evidentially-informed acceptance of one treatment and rejection of alternatives, and researchers cannot be overly dogmatic in their espousal of new treatments. The principle tends towards caution when deciding what research should be conducted.

Nevertheless, although this principle is of considerable importance, it is debated (London 2009; Gelfand 2013), particularly because it conflates the different ethical approaches of research and practice, between the physician and the researcher (Miller and Brody 2003; Veatch 2007; Miller and Joffe 2011). This tension is found at critical moments in medical research, as practitioners are

under a duty to cure or heal, acting in the patients' best interests, whereas researchers may take a more distanced position, in establishing the effectiveness of different techniques or pharmaceuticals. There is a mismatch between the care of an individual patient and wider long-term gains for the population as a whole. For example, given the severity of the recent Ebola outbreak in West Africa it was considered by some (e.g. Adebamowo et al. 2014) to be unethical to establish a control group when testing new drugs, as all patients should have received any untested new treatments. Even though there was still genuine uncertainty about the new treatment, it was considered wrong to withhold it from anyone. Further, the degree of uncertainty may be disputed given the existing evidence, because it could be considered malpractice to carry out the new treatment if there is not genuine uncertainty, and what constitutes the boundary between certainty and uncertainty within medical knowledge becomes crucial, which in turn depends on the rigour of existing research: one researcher might think that the current research is inadequate or flawed and therefore consider a new treatment to be worthy of testing, whereas another might consider current research to be rigorous and dependable, and therefore the testing of a new treatment is inappropriate (e.g. Mirzadeh et al. 2014, Mirzadeh and Ponce 2015; Montgomery 2015).

The principle of equipoise is not unknown in educational research. Gorard (2013) discusses the principle, particularly arguing for the advantages of a 'waiting list design' (p. 134), in which a treatment is deferred for the initial control group, so that they can initially be compared to the intervention groups, but then get the benefit of its potential effects. Fives et al. (2015) also describe its benefits, suggesting that:

> ...a reasonable distinction can be made between what is taken as best practice in education on the one hand and on the other hand the social scientist's honest null hypothesis when considering whether an innovative programme will prove to be more effective than an alternative (p. 60)

In their research, the expert *education* community considered that a particular reading intervention was effective but the expert *research* community considered that there was no valid evidence to support this claim. In medical research, some would argue that the opinion of the research community should prevail, as equipoise is an illusion – the 'difference' position (e.g. Miller and Brody 2003); Fives et al. would disagree, and considered that through negotiation the research design should be modified to ensure that participants were not deprived of the opportunity to benefit from the treatment. Their solution involves a wider discussion between the different interested communities involved, with their different ethical priorities.

There was however a further complexity with such an approach in the CTG project, as, first, the schools were increasingly involved in aspects of the research and encouraged to conduct their own research, acting both as researchers and as practitioners, and second, they were also deemed to be the givers of consent on behalf of the pupils. They therefore would have had to reconcile three potentially different roles: researchers, practitioners, and consent givers. As researchers, there might be good reasons for conducting a trial of a particular intervention, and as practitioners, there might be a general professional belief in the efficacy of the intervention, but as consent givers, the particular intervention might not suit an individual pupil for personal reasons, such as major anxieties caused by any process of testing. Within a school, it might be that different staff took on these different positions, but this would of itself need leadership

decision about who should decide on these different priorities, and how any conflict would be resolved.

Despite its importance, the notion of equipoise has received little sustained attention in some professional discussions of ethics in educational research. Neither BERA (2011) guidelines nor the American Educational Research Association's (2011) code of ethics use the term at all. BERA do mention some ethical issues raised by RCTs, touching on the question of beneficence:

> Researchers must take steps to minimize the effects of designs that advantage or are perceived to advantage one group of participants over others e.g. in an experimental or quasi-experimental study in which the treatment is viewed as a desirable intervention and which by definition is not available to the control or comparison group respectively. (BERA 2011, p. 7)

This partially addresses the problem, for instance by providing the 'treatment' to the control group after the research is completed, but arguably does not go far enough in that simply *minimising* the effects is insufficient. The risks should not exist at all, rather than merely being reduced. For example, in a comparison of literacy strategies, once a group of pupils has been taught using one approach which is shown to be ineffective, it would be difficult and arguably more detrimental, to start again with another approach. Equipoise demands consideration of whether a treatment should run at all, rather than how best to alleviate any subsequent negative effects after the event.

This lack of attention in BERA's guidance may also account for its absence in many current guides to the design of RCTs in education. For instance, The Educational Endowment Foundation, which seeks to promote greater use of and reliance on RCTs in education, whilst setting out its ethical principles, makes no mention of equipoise (EEF 2015), although this could be a valuable deliberative tool for potential researchers to balance the potential competing principles.

The second stage is that of the use of interim data. An implicit equipoise question arose: was there still 'genuine uncertainty' about a particular strategy if the interim data showed that it was currently ineffective? Hypothetically, if the analysis had shown that one of the interventions was having significantly negative effects on pupils learning, the researchers could have cancelled that intervention, as discussed above. In medical research, 'knowledge of short-term outcomes may directly and legitimately affect the decision to continue or abandon a...trial' (Allardyce et al. 2012. p. 126). This might include a statistically insignificant number of patient deaths, if this is seen as too high (see Pocock 1992), though current practice is often for the researchers to tend to keep interim data confidential from the clinicians involved (Allardyce et al. 2012). The question of defining genuine uncertainty can be seen as both an ethical and an epistemological one, for instance in terms of the ethical use of statistical analysis (Sammonds 1989), for instance, setting the appropriate p-values for the interim data, or in terms of reliance on other forms of evidence and research (Lie and Miller 2012).

Within the CtG project, the researchers' process of ethical deliberation would have benefitted from applying this principle. This is not to suggest that the research was therefore unethical, but that the principle would have allowed for the tensions between the question of beneficence/non-malfeasance and that of certainty/uncertainty to be made more explicit, and thus aid the decision-making process. It is hard to say if the interventions that were finally chosen would have been rejected, or that different decisions on the basis of the interim data would have been made, but problem was that at these stages, the researchers did not explicitly identify any rules or guidelines for deliberating these issues.

In future school-based RCTs, the problem of the handing of interim date would potentially be played out through the triple roles identified above. Thus, it raises further questions for schools are researchers, as expert practitioners and potentially as givers of consent. It is not hard to imagine that interim research data on the proven efficacy of an intervention might lead to demands for its expansion as school practice, and to the abandonment of a control group, or vice versa if the intervention was shown to be harmful or less successful than the control treatment.

CONCLUSION

In conclusion, the initial impulse for the research, to raise the attainment of a type of pupil who has often been let down by the process of schooling, should be re-acknowledged. RCTs have a vital role - alongside other methodologies - to play in contributing to our understanding of how we might address these issues. This article has sought to reflect creatively on the ethical deliberations within one uniquely large research project by identifying the various ethical issues it raised, and how some of them might have been addressed more effectively. These issues themselves are probably not unique, potentially applying to other RCTs, and thus offer the opportunity for other researchers to respond imaginatively to the ethical complexities involved.

In particular, equipoise is reasserted as an important conceptual tool for thinking through the dilemmas, defined as genuine uncertainty within the expert community about a particular strategy, supporting the arguments of Gorard (2013) and Fives et al (2013). It would need to be nuanced considerably in determining the nature and limits of genuine uncertainty as well as the constituency of an expert community. These are questions to be developed and refined in due course, in relation to particular research. This principle is also important because of the ways that it appears twice in the research process, first at the design stage and then once there is any interim data. This means that it is not an issue that can be simply addressed in the initial stages, but is one which researchers need to be ready to address throughout a project. In particular its relationship with consent would benefit from more elaboration, especially in relation to who makes these judgements.

Finally, the ethical issues raised by the development of school-led RCTs also need further consideration, as the different roles of teacher and researcher, and potentially consent-giver, also

need to be addressed. If teachers are to be mobilised to tackle educational inequality by conducting RCTs, then the principle of equipoise provides the start of a deliberative discussion about these issues, but cannot be applied uncritically in seeking to make judgements about the tangle of rights, responsibilities, benefits and uncertainties at stake.

REFERENCES

Adebamowo C., O. Bah-Sow, F. Binka, R. Bruzzone, A. Caplan, J. Delfraissy, D. Heymann P. Horby, P. Kaleebu, J.-J.Muyembe Tamfum, P. Olliaro, P. Piot, A. Tejan-Cole, O. Tomori, A. Toure, E. Torreele and J. Whitehead. 2014. "Randomised controlled trials for Ebola: practical and ethical issues". *Lancet* 384: 1423-4.

Allardyce, R., P. Bagshaw, C. Frampton, F. Frizelle, P. Hewett, P. McMurrick, N. Rieger, J. Smith, M. Solomon, and A. Stevenson. 2012. "Ethical Issues with the Disclosure of Surgical Trial Short-term Data". *ANZ Journal of Surgery*. 81(3): 125-131.

Bentham, J. 1996. *An Introduction to the Principles of Morals and Legislation*. Oxford: Oxford University Press.

British Educational Research Association. 2011. *Ethical Guidelines for Educational Research*. BERA: London. [https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf?noredirect=1]

Burgess, R. (ed.) 1989. *The ethics of educational research*. Falmer: London.

Chalmers, D. 2011. "Viewpoint: Are the research ethics committees working in the best interests of participants in an increasingly globalised research environment?" *Journal of Internal Medicine*, 269(4): 392- 395.

Chard, J. and R. Lilford. 1998. The use of equipoise in clinical trials, *Social Science & Medicine*, 47(7): 891-898.

Churches, R. 2016a. *Closing the Gap: Test and Learn. Executive summary.* National College for Teaching and Leadership.

Churches R. 2016b. *Closing the gap: test and learn. Research report.* National College for Teaching and Leadership.

Churches, R. and T. McAleavy. 2015. *Evidence that counts – what happens when teachers apply scientific methods to their practice.* Reading: CfBT.

Educational Endowment Foundation, 2015. *EEF Ethics Policy*. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/EEF_Ethics_Policy_2015.pdf

Fives, A., D. Russell, J. Canavan, R. Lyons, P. Eaton, C. Devaney, N. Kearns and A.O'Brien 2015. "The ethics of randomized controlled trials in social settings: can social trials be scientifically promising and must there be equipoise?" *International Journal of Research & Method in Education*, 38(1): 56-71.

Frankfort-Nachmias, C. and Nachmias, D. 1992. *Research methods in the social sciences. London: Edward Arnold.*

Freedman, B. 1987 "Equipoise and the ethics of clinical research". *The New England Journal of Medicine*, 317(3): 141–145.

Fried, C. 1974. *Medical Experimentation: Personal Integrity and Social Policy*. Amsterdam, The Netherlands: North Holland Press.

Fries, J. and E. Krishnan. 2004. "Equipoise, design bias, and randomized controlled trials: the elusive ethics of new drug development". *Arthritis Research and Therapy*. 6(3): R250–R255.

Gelfand, S. 2013. "Clinical equipoise: actual or hypothetical disagreement?" *The Journal of medicine and philosophy*, 38(6): 590–604.

Goldacre, B. 2013. *Building evidence into education*. London: Department for Education.

Gorard, S. 2013. *Research design: Creating robust approaches for the social sciences*. London: Sage.

Haggerty, K. 2004. "Ethics Creep: Governing Social Science Research in the Name of Ethics". *Qualitative Sociology* 27: 391 -414.

Hammersley, M. 2008. "Paradigm war revived? On the diagnosis of resistance to randomized controlled trials and systematic review in education". *International Journal of Research & Method in Education*, 31(1): 3-10.

Hammersley, M. 2010. "Creeping ethical regulation and the strangling of research". *Sociological Research Online*, 15(4): 16.

Haynes, L., O. Service, B. Goldacre and D. Torgerson. 2012. *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials.* London: Cabinet Office.

Homan, R. 2002. "The principle of Assumed Consent: the Ethics of Gatekeeping". In M. McNamee and D. Bridges (eds.). *The Ethics of Educational Research*. pp. 23-40. Oxford: Blackwell.

Israel, M. & I. Hay. 2006. *Research Ethics for Social Scientists: Between Ethical Conduct and Regulatory Compliance*. London: Sage.

James, M. 2013. "New (or not new) directions in evidence-based practice in education". London: BERA. Downloaded 16/02/2016 from: https://www.bera.ac.uk/promoting-educational-research/issues/dfe-review-of-evidence-in-education

Kaplan, I., S. Miles and A. Howes 2011. "Images and the ethics of inclusion and exclusion: learning through participatory photography in education". *Journal of Research in Special Educational Needs* 11(3): 195–202.

Kim S. and F. Miller. 2015. "Varieties of Standard-of-Care Treatment Randomized Trials: Ethical Implications". *Journal of the American Medical Association*. 313(9): 895-896.

Lie, R. and F. Miller. 2011. "What counts as reliable evidence for public health policy: the case of circumcision for preventing HIV infection". *BMC Medical Research Methodology* 11: 34.

Lilford, R. and J. Jackson. 1995 "Equipoise and randomisation". *Journal of the Royal Society of Medicine* 88: 552 -559.

Locke T., N. Alcorn and J. O'Neill 2013. "Ethical issues in collaborative action research". *Educational Action Research* 21(1): 107-123.

London, A. 2009. “Clinical Equipoise: Foundational Requirement or Fundamental Error?” In B. Steinbock (ed.) *The Oxford Handbook of Bioethics.* Retrieved 12 Nov. 2015, from http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199562411.001.0001/oxfordhb-9780199562411-e-025

McNamee, M. and D. Bridges (eds.) 2002. *The Ethics of Educational Research*. Oxford: Blackwell.

Midgley, W., P Danaher and M. Baguley (eds.) 2013. *The Role of Participants in Education Research: Ethics, Epistemologies, and Methods.* Abingdon: Routledge.

Miller, F. G. and H. Brody. 2003. “A Critique of Clinical Equipoise: Therapeutic Misconception in the Ethics of Clinical Trials”. *The Hastings Center Report*, 33(3): 19–28.

Miller, F. and S. Joffe. 2011. “Equipoise and the Dilemma of Randomized Clinical Trials”. *New England Journal of Medicine*, 364(5): 476-480.

Mirzadeh Z, K. Chapple, M. Lambert , R. Dhall, and F. Ponce 2014. “Validation of CT-MRI fusion for intraoperative assessment of stereotactic accuracy in DBS surgery”. *Movement Disorders*; 29: 1788–1795.

Mirzadeh, Z. and F. Ponce 2015. "DBS with versus without MER: clinical equipoise or malpractice?" *Movement Disorders*. 30(3): 439-441.

Mockler, N. 2014. "When 'research ethics' becomes 'everyday ethics': the intersection of inquiry and practice in practitioner research". *Educational Action Research* 22(2): 146-158.

Montgomery, E. 2015. "Validation of CT-MRI fusion for intraoperative assessment of stereotactic accuracy in DBS surgery". *Movement Disorders*, 30(3): 439.

Morrison, K. 2001. "Randomised Controlled Trials for Evidence-Based Education: Some Problems in Judging 'What Works'". *Evaluation and Research in Education* 15(2): 69–83.

National College for Teaching and Leadership 2016a. *Closing the gap: test and learn. Technical annex A: statistical analysis.*

National College for Teaching and Leadership 2016b. *Closing the gap: test and learn. Technical annex B: provider perspectives.*

National College for Teaching and Leadership 2016c. *Closing the gap: test and learn. Teacher led randomised controlled trials - Digital applications.*

National College for Teaching and Leadership 2016d. *Closing the gap: test and learn. Teacher led randomised controlled trials – Feedback case studies.*

National College for Teaching and Leadership 2016e. *Closing the gap: test and learn. Teacher led randomised controlled trials – Literacy case studies*.

National College for Teaching and Leadership 2016f. *Closing the gap: test and learn. Teacher led randomised controlled trials – Numeracy case studies*.

National College for Teaching and Leadership 2016g. *Closing the gap: test and learn. Teacher led randomised controlled trials – Organisation case studies*.

National College for Teaching and Leadership 2016h. *Closing the gap: test and learn. Teacher led randomised controlled trials – Resilience case studies*.

Parsons, S., C. Abbott, L. McKnight, and C. Davies 2015. "High risk yet invisible: conflicting narratives on social research involving children and young people, and the role of research ethics committees". *British Educational Research Journal,* 41(4): 709–729.

Pocock, S. 1992. "When To Stop A Clinical Trial". *BMJ: British Medical Journal*, 305(6847): 235–240.

Robinson-Pant, A. and N. Singal 2013. "Researching ethically across cultures: issues of knowledge, power and voice". *Compare: A Journal of Comparative and International Education*, 43(4): 417-421.

Sammons, P. 1989. “Ethical issues and statistical work”. In R. Burgess (ed.) *The ethics of educational research*. Pp. 31-59. Falmer: London.

Shamim F. And R. Qureshi 2013. “Informed consent in educational research in the South: tensions and accommodations”. *Compare: A Journal of Comparative and International Education*, 43(4): 464-482.

Taber, K. 2013. The right medicine for educational research? Royal Society of Chemistry. Downloaded on 16/02/2016, from: http://www.rsc.org/Education/EiC/issues/2013May/goldacre-education-research-report.asp

UNICEF (1989) *The United Nations Convention on the Rights of the Child*. London: UNICEF. http://www.unicef.org.uk/Documents/Publication-pdfs/UNCRC_PRESS200910web.pdf

Veatch, R. 2007. “The Irrelevance of Equipoise”. *Journal of Medical Philosophy*, 32(2): 167–183.

Walford, G. 2005. “Research Ethics Guidelines and Anonymity.” *International Journal of Research and Method in Education* 28(1): 83–93.

Webster, T. 2013. “Doing What Works: Challenges to being Ethically ‘Reasonable’”. In Midgley, W., P. Danaher and M. Baguley (eds.) *The Role of Participants in Education Research: Ethics, Epistemologies, and Methods.* pp 64-81. Abingdon: Routledge.

Zeni, J. 2001. *Ethical issues in Practitioner research*. New York: Teachers College Press.