Iris Eireannach An
OIDEACHAIS
The Irish Journal of
EDUCATION

Randomised Controlled Trials in Education: An Experiment Worth Taking?

John O'Connor

Qualifications and Quality Assurance Authority of Ireland

Abstract

Appropriately designed and implemented randomised controlled trials (RCTs) are increasingly promoted as the gold standard for objective evaluations of effectiveness and impact in education. The rise in the use of RCTs in education in many countries is based on the belief that by determining causal relationships between interventions and outcomes of interest, impartial measures of the effectiveness of specific policies and classroom practices can be established. There are longstanding concerns about the use of RCTs in the social sciences, and both the logic and the appropriateness of RCTs in education are hotly contested in the international literature. Not so in Ireland, however, where the topic has attracted little or no critical attention, and where there is no tradition of conducting large-scale RCTs in education. This article examines the key features of RCTs and their epistemological and ontological basis, situating the current wave of international interest in RCTs in education within a wider historical context. It then presents some of the main critiques and criticisms of RCTs in education before inviting educationalists with an interest in the use of educational research in policy and practice in Ireland to critically engage with the opportunities and challenges presented by RCTs.

Keywords: randomised controlled trials, education, Ireland, research, policy, practice

Author Note

John O'Connor may be contacted at joconnor@qqi.ie

John O'Connor's ORCID iD is: https://orcid.org/0000-0001-9027-8214

The author received no financial support for the research, authorship, and/or publication of this article.

Please cite as: O'Connor, J. (2024). Randomised controlled trials in education: An experiment worth taking? *Irish Journal of Education*, *48*(4), 1-27. www.erc.ie/ije

The centrality of education to all our lives means that the public have a legitimate interest in the effectiveness of educational policy and practice and a low tolerance for any waste in public expenditure. Systematic use of the best available research findings in education can improve the quality and transparency of decision-making and protect students from potentially harmful effects of fads and quick-fix solutions, or prevailing political ideologies (Darling-Hammond et al., 2020; Hargreaves, 1996; OECD, 2023a; UNESCO, 2023). Since the 1990s, in Ireland and elsewhere, there has been growing academic interest in the creation, mobilisation, and use of research evidence in education (Biesta, 2007; Davies, 1999; Gleeson, 2012; O'Connor, 2024; Whitty, 2006). Many governments have embraced the rhetoric of evidence-based policy (EBP) often inspired by the "what works" movement in education (Brown & Malin, 2022; Cairney, 2016; OECD, 2022, 2023c; Pellegrini & Vivanet, 2021; Slavin, 2020). The prospect of proven solutions to complex educational problems has an obvious appeal for busy policymakers and school leaders. Notwithstanding the rhetoric of EBP in education, progress has been slow. Many decisions about improving teaching and learning methods, reform of curriculum, or efforts to combat educational disadvantage are arrived at without systematic engagement with available scientific research and in the absence of impartial and robust evidence about effectiveness (Brown & Malin, 2022; Gleeson, 2012; Pegram et al., 2022).

Commenting on schools in the UK, Ben Goldacre (2013) warned that "right now, there are huge numbers of different interventions used throughout the country – different strategies to reduce absenteeism, or teach arithmetic, or reduce teenage pregnancies, or any number of other things – where there is no evidence to say which of the currently used methods is best" (p. 12). For Goldacre, the solution was the experimental trial, which he believed could "replace the current system where untested methods are passed to teachers through a variety of often dubious outlets" (p. 12). Also commenting on the UK, Styles and Torgerson (2018) concluded that too many initiatives are rolled out in schools, without rigorous pre-trialling, or in circumstances where evaluation is postponed or never done, because that would delay getting a programme off the ground. Would these criticisms be unfamiliar to policymakers and educational researchers in Ireland, or are they confined to our nearest neighbour? Most, if not all countries, including Ireland, recognise the importance of robust evaluation studies in education and the imperative for ongoing systematic and critical engagement by policymakers and practitioners with valid research evidence (Darling-Hammond et al., 2020; OECD, 2022, 2023a).

What counts as valid research evidence in education remains a vexed question and a central issue for this paper. Classical experimentalism is seen by many as the most rigorous and impartial form of educational research to determine "what works" in education and randomised controlled trials (RCTs) are the current poster child of the experimental approach (Hammersley, 2003; Hutchison & Styles, 2010; Pampaka et al., 2016; Slavin, 2002; Torgerson & Torgerson, 2008). This article examines the key

features of RCTs and their epistemological and ontological basis, situating the current wave of international interest in RCTs in education within a wider historical context. Some of the main critiques and criticisms of RCTs in education are also presented before inviting educationalists with an interest in the use of educational research in policy and practice in Ireland to critically engage with the opportunities and challenges presented by RCTs.

RCTs – A Primer

Experimental and Control Groups, and Outcome Measures

RCTs have unique design features and involve statistical operations that underpin their knowledge claims. Interested readers can find many excellent and detailed methodological guidelines for conducting RCTs – see, for example, Torgerson and Torgerson (2008) or Hutchison and Styles (2010). For a detailed examination of the statistical underpinnings of RCTs, see Coe (2002). Typically, RCTs are used either as prospective studies to pre-test new policies or programmes or as evaluations of an existing programme, including comparing two or more programmes that are intended to influence the same outcome (Haynes et al., 2012; Torgerson & Torgerson, 2008). RCTs are designed to compare outcomes of interest for a group of individuals who receive an intervention (the experimental group) with the outcomes of a comparable group which does not receive the intervention (the control group). Valid outcomes used as a measure of effectiveness are identified and declared in a study protocol before the experiment is conducted. To minimise the risk of bias exaggerating treatment effects, careful consideration is given to the selection of outcome measures, how outcomes are measured, and by whom. Standardised test scores are often used as a cost-effective and independent outcome measure in RCTs.

Randomisation

The unit in an RCT in education can be an individual student, a class, or a school. Units are randomly assigned to either the experimental group or the control group so that each student, class, or school has an equal or random chance of receiving an intervention. Randomisation is intended to minimise selection bias by ensuring that the allocation of participants to groups is independent from the measurement of outcomes of interest (Oliver et al., 2010; Torgerson & Torgerson, 2008). The allocation of participants to experimental or control groups can be based on simple randomisation, such as the toss of a coin, or on the basis of restricted randomisation to ensure that certain characteristics, such as school size, are evenly distributed across groups (Torgerson & Torgerson, 2013). Individual participants can be grouped or "clustered" so that a class or school is treated as a single observable unit and randomly

allocated accordingly. Clustering is often practical in a natural teaching situation, but comparing and interpreting cluster-level mean test scores of classes or schools must be done with caution to avoid biased estimates of statistical significance (Spybrook et al., 2020; Torgerson & Torgerson, 2013).

Randomisation is what distinguishes RCTs from quasi-experimental and non-experimental methods and is central to the claim that an outcome of interest is not likely to be the result of some variable other than the intervention under investigation. Experimental results that isolate cause-and-effect in an RCT can be generalised to a wider population of interest, i.e., all those who will receive the intervention if selected but this is subject to a high level of similarity in the characteristics of the experimental sample and the target population (Connolly et al., 2018; Hutchison & Styles, 2010). Despite the benefits of randomisation, generalising experimental findings about the effectiveness of educational interventions remains a key question of interest to researchers, policymakers, and practitioners.

Effect size

Coe (2002) defines effect size as "simply a way of quantifying the size of the difference between two groups. It is easy to calculate, readily understood, and can be applied to any measured outcome in education or social science" (p. 1). A common way of standardising effect sizes is by dividing the difference between the mean scores of the experimental and control groups by the standard deviation. In an experimental setting, for example, where test results of an experimental group and a control group are compared, using the Cohen's *d* as one of the most frequently used effect size measures, an effect size of 0.6 would mean that the score of the average person in the experimental group would be 0.6 standard deviations above the score of the average person in the control group. Advocates of effect sizes in education find that they are more influential on time-poor policymakers and educators when communicated in familiar and meaningful terms, such as school grades, months of education, or percentile rank improvement scores (Baird & Pane, 2019). In the UK, the Education Endowment Foundation (EEF) translate effect sizes into "additional months of learning", with the current education and learning toolkit reporting on 30 educational interventions which show effect sizes ranging from +7 months (metacognition and self-regulation) to –3 months (repeating a year) (see https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit).

If a single RCT can generate impartial evidence about effectiveness, then meta-analysis, the synthesis of the results of multiple RCT studies into an overall average effect size, can produce cumulative differences in results between experimental and control groups. The meta-analysis of results from replicated RCT studies across a wide range of educational settings has been described as the gold standard for educational research concerned with effectiveness and the generalisability of effects (Bennett,

2013; Darling-Hammond et al., 2020; Hattie, 2023; Pegram et al., 2022; Slavin, 2020; Van Damme, 2022).

According to Cohen (1988), reported effect sizes can be "small", "medium", or "large". Meta-analyses of RCTs in education often report "small" or "medium" effect sizes for interventions (Coe, 2002; Hedges & Schauer, 2018), simply reflecting the reality that there are no silver bullets or quick-fix solutions in education. Statisticians argue that small effect sizes can be very significant over time, especially where all students benefit and where the costs of implementation are relatively low (Coe, 2002). Arguably, RCTs that demonstrate zero beneficial effects of an educational policy or intervention are just as useful for decision-makers. The practice of registering RCTs and publishing their results regardless of the reported effect size is something that most researchers would support given the prevalence of publication bias where negative or null results routinely go unreported or unpublished (Kicinski, 2014; Ropovik et al., 2021).

Randomisation, effect sizes, and meta-analysis are central methodological features of RCTs, which involve complex statistical processes and operations and require expert statistical knowledge and guidance. Even so, many educational researchers remain dubious about the statistical power of RCT findings, and the persuasive power of numbers, and argue that questions about what counts as scientific knowledge in education must be placed within a wider philosophical context of educational research, to which we now turn.

Epistemological and Ontological Basis of "What Works" in Education

The epistemological paradigm underpinning the "what works" thinking and the privileging of experimental studies in education is associated with *empiricism* – the position that valid knowledge is based on human experience gained through the senses and accordingly we find out about the world by observation and experimentation; and *positivism* – a strong form of empiricism where factual knowledge is seen as gained through quantifiable observations and statistical analysis where the researcher acts as an independent data analyst (Ball, 2015; Biesta, 2007; Whitty, 2006).

RCTs are a suitable method within a positivist ontology and epistemology and are well-placed to address causal questions about "what works" in education (Slavin, 2020). RCTs rely on deductive reasoning to determine causal relationships and test for their generalisability. Supporters of classic experimental methods claim they can produce law-like regularities about human behaviour and "assurances of unambiguous and accurate knowledge of the world" (Crotty, 1998, p. 18). This belief in determinism, where behaviours have identifiable and measurable causes, leads to bold claims about warranted assertions. Lunn and Robertson (2018) claim that "the power of the experimental method derives from its ability to identify causal effects. Where outcomes

are compared under two conditions that differ by a single factor, we can be confident that any difference observed is due to that single factor" (p. 8). Leading positivistic educational researchers such as Robert Slavin (2002, 2008a, 2008b, 2020) claim that scientific evidence about and for education could only result from studies and synthesis of studies that estimate the magnitude of the effect of an educational intervention or programme on outcomes of interest. Slavin anticipated that the widespread use of controlled experiments in education would "transform the practice of education as well as research in education and (…) set education on the path toward the kind of progressive improvement that most successful parts of the economy and society embarked on years ago" (Slavin, 2002, p. 20). Almost twenty years on, Slavin came to the view that evidence of effectiveness could only come from "rigorous experiments in which students experiencing experimental programs are compared over significant periods (say, a semester or more) to those using traditional control methods in terms of gains on valid measures of achievement or other outcomes" (Slavin, 2020, p. 22). The track record of the experimental method in the natural sciences has been used to promote the use of RCTs in education to busy policymakers enticed by the prospect of research findings that can guide binary decisions, such as to fund or not to fund a programme.

The current wave of enthusiasm for experimental methods in education has not always been the case. For most of the past century, support for the methodological orthodoxy of experiments as the one best way to study the effectiveness of policy and practice in education has fluctuated with the rise and fall of competing epistemological and ontological paradigms (Connolly et al., 2018; Hedges & Schauer, 2018; White, 2019). A brief account of some of these key paradigmatic shifts provides some historical context to the trajectory of RCTs in educational research.

Positivism and postpositivism

The mid-twentieth century popularity of experimental approaches in education reflected the dominant positivist paradigm at the time, which assumed that the logical and mathematical treatment of data derived from sensory experiences constituted the one true way of knowing (Phillips & Burbules, 2000). The prevailing view was captured in this account from Campbell and Stanley (1963), who regarded the experiment in education as:

> the only means for settling disputes regarding educational practice, as the only way of verifying educational improvements, and as the only way of establishing a cumulative tradition in which improvements can be introduced without the danger of a faddish discard of old wisdom in favour of inferior novelties (p. 2)

At that time, non-experimental forms of educational research were seen as lacking

scientific rigour and in need of reform. In 1943, Fred Clarke, then director of the Institute of Education in the UK from 1935 to 1945 and the first chair of the National Foundation for Educational Research in the UK, criticised educational research for its "slipshod carelessness" … "disregard for precision of thought" … and "reckless play of sentimentality" (quoted in Armytage, 1976, p. 11).

1970s – Bonfire of the Certainties and the Paradigm Wars

The late 1970s saw the mass rejection of the western world's scientific utilitarian culture and the "empirical turn" away from positivism. Researchers in the social sciences, including education, began to offer more realist and interpretive accounts of the social world. They rejected mechanical models of causality in favour of heterogeneity, subjectivity, multiplicity, complexity, and irreducibility as they sought to explain the "post-modern condition" of education (Biesta, 2007; Hammersley, 2003). For interpretivists, individual and subjective accounts of social reality came to matter more than correlations, statistics, or measures of significance, and educational researchers began to challenge impartiality of researchers and objective facts as imperatives for scientific knowledge (Ball, 1997; Feuer et al., 2002; Lather, 2008). Borrowing ideas from the natural sciences for the study of education was seen by interpretivists as deeply flawed. The idea of impartiality was seen as creating an unrealistic impression of educational researchers placed in a neutral and ambivalent position in relation to educational phenomena that have "causal explanations brought about by active independent variables operating on passive dependent variables" (Hudson, 1975, p. 69).

Cronbach (1975) and others had sparked a growing pessimism in the social sciences about making inferences from the particular to the general. Population generalisability became a contested condition for scientific educational research, in simple terms "similar behaviour has no necessary implications for similarity of reasons for that behaviour" (Cope, 1975, p. 71). Many social scientists argued that research findings had been interpreted within local contexts and these tensions around objectivity and generalisability came to a head in the "paradigm wars" of the 1980s. There followed a significant downturn in the number of RCTs conducted internationally and particularly in the US where that period was seen as "a dark era for those interested in randomised trials in education" (Hedges & Schauer, 2018, p. 270).

The Latest Turn

While the first RCTs in education were conducted as far back as the 1920s, pre-dating their use in medical research, their incidence has increased rapidly over the past three decades (Connolly et al., 2018; Hedges & Schauer, 2018; Hutchison & Styles, 2010; Styles & Torgerson, 2018). Following a decline in the number of RCTs conducted in

education during the 1970s and 1980s, the mid-1990s brought renewed political interest in the economic and social benefits of education, encouraged by supranational organisations calling for educational research to address real-world problems and directly contribute to policies and practice that improved student outcomes (European Commission, 2007; OECD, 2007). Questions about "what causes people to become educated" were once again accompanied by the authoritarian "methods that work project" (Clark, 2005, p. 289). The pursuit of school effectiveness and school improvement agendas was associated with a more technical rationalist framework and seen by some as legitimising the governance of education using findings from experimental methods (Ozga, 2012; Rowe & Oltmann, 2016).

While the orthodoxy of the experimental methods in education has ebbed and flowed since the middle of the last century, international evidence suggests that interest in "what works" is here to stay and it is challenging and changing the nature of educational research (Cochran-Smith & Lytle, 2006; Hedges & Schauer, 2018; OECD, 2022).

International Experience of RCTs

The resurgence of interest in RCTs in education has been particularly marked in the US, the UK, New Zealand, and across northern European countries, where apathy and antipathy towards RCTs as an empirical method during the paradigm wars has given way to strong demand for proven educational interventions based on RCTs (Styles & Torgerson, 2018).

In the US, the course of educational research changed dramatically in the early 2000s (Hedges & Schauer, 2018). The No Child Left Behind Act in 2001 and the Education Sciences Reform Act 2002 made government funding for educational research contingent on RCTs. In response, the Institute of Education Science was established, which, in turn, set up the What Works Clearinghouse (WWC) to set standards for publicly funded educational research in the US. The WWC database now provides extensive resources for educators interested in the effectiveness of a wide range of interventions from homework to performance-related remuneration for teachers (https://ies.ed.gov/ncee/wwc/). The highest WWC rating for rigour is reserved for evaluations of interventions based on RCT results.

Partly inspired by developments in the US, the EEF in England was established in 2011 and funded by the Department for Education to use evidence to break the link between family income and educational achievement. As part of the UK "What Works Network", the EEF funds large-scale school-based evaluations, mostly based on individual RCTs, and synthesis studies of multiple RCTs. Results are disseminated in the online *Teaching and Learning Toolkit* (https://educationendowmentfoundation.org.uk/), which offers teachers useful information about what works in the classroom based on existing evidence.

The New Zealand Ministry of Education established the *Best Evidence Synthesis* (https://www.educationcounts.govt.nz/topics/bes), a programme of commissioned syntheses of RCT effect sizes, aimed at identifying and explaining the interventions that make a bigger difference than "business as usual" on valued outcomes for all students.

It is clear from these examples that governments play a leading role in legitimising and commissioning large-scale experimental studies in education. Governments are well-placed to encourage schools to adopt what are deemed to be effective programmes (Slavin, 2020), but this means that schools are incentivised to participate in RCTs, and this cannot be taken for granted. Insights into the work of the EEF in England suggest that keeping schools centrally involved in planning at all stages of an RCT is crucial to overcoming any resistance. Schools in England, initially unfamiliar with RCTs and what was involved, were concerned that trials would deprive some students from promising interventions. Over time, schools became acquainted with the work of the foundation and began to see the benefits of RCTs for education (Edovald & Nevill, 2020).

The synthesis of RCT findings in education is not only undertaken by governments or their agencies. Hattie (2009, 2023) synthesised existing meta-analyses of experimental and quasi-experimental studies. His most recent work involved 2,100 meta-analyses drawn from over 13,000 studies based on the participation of more than 400 million students. Hattie's findings are presented as a compendium of proven initiatives for the improvement of student achievement. Though still influential on teaching practice, the large effect sizes reported by Hattie in his original synthesis have been challenged (Bergeron & Rivard, 2017; Blass, 2020; Terhart, 2011; Wrigley & McCusker, 2019). Because Hattie aggregated effect sizes without considering important differences in the context of individual studies, and because he mixed randomised and non-randomised studies within his synthesis, reported effect sizes must be treated with caution. Whether undertaken by government agencies or independent researchers, meta-analysis is a complex research method. When employed in educational settings, it should be under the guidance of expert statisticians, and results, expressed as effect sizes, must be critically and carefully interpreted by educationalists. Beyond the specific statistical concerns referenced above about the synthesis work of Hattie, there are more general criticisms and critiques of RCTs that cannot be ignored.

RCTs – Criticisms, Critiques, and Limitations

The use of RCTs in education has been met with opposition from different perspectives. Some question the validity of experimental findings in education and argue that RCTs are simply unsuited to the study of complex social processes such as education. Others see that the uptake of RCTs is part of a wider neoliberal plot by policymakers to exercise ever-greater control of education, students, teachers, and research. RCTs, like

all research methods, have their limitations and researchers and policymakers should maintain a sceptical outlook in reaching conclusions about their contribution to valid knowledge.

An Interpretivist Take

The prevailing discourse around EBP privileges experimental and quantitative findings in the belief that it is possible to ascertain objective versions of reality and universal truths about what works in education (Cartwright, 2019; Gough, 2021; White, 2019). Positivists and post-positivists adhere to a deterministic philosophy where causes determine effects or outcomes (Creswell, 2014). In contrast, interpretivists see schools as complex social systems unsuited to the determination of discrete cause-and-effect relationships. For interpretivists, RCTs in education are crude technical instruments, stripped of values and context which are developed and applied "without scholarship" and with a fetish for statistical analysis that reduces the method to "mere abstracted empiricism" (Siraj-Blatchford & Siraj-Blatchford, 1997, p. 243). For interpretivists, attempts to make sense of educational phenomena by neutralising the practice and study of education through the façade of objectivity are rejected as naïve and unproductive, as Stephen Ball (1997) writes: "the idea that human sciences like educational studies…somehow have a neutral status embodied in a free-flowing progressive rationalism are dangerous and debilitating conceits" (p. 264). A critical sociologist of education, Ball has been critical of the hegemony of EBP, and of the privileging of hyper-rational techniques such as RCTs, which he sees as legitimising greater control of research by policymakers. For Ball, experimental methods in education assume that education is a sterile process, and he dismisses effect sizes as "accounts of what works for unselfconscious classroom drones to implement" that threaten to impose "an absolute standardisation of research purposes, procedures, reporting and dissemination" (Ball, 2001, pp. 266-267). Writing in the same critical tradition, Gert Biesta (2007, 2010, 2015), dismisses the positivist paradigm in education research for failing to take account of the complexity of education as a social practice, arguing that "despite the attempts of many to transform education into a causal technology, the simple fact is that education is not a process of push and pull … [it] is an open and recursive system" (2007, p. 8). Biesta went on to call for a fundamental and urgent rethink of the EBP project in education, recognising the centrality of values and the normative dimension of education and educational research and the failure of experimental research to take account of social interaction and power relations in education (Biesta, 2010).

Ethical, Methodological, and Practical Limitations of RCTs

Critics of RCTs in education on philosophical grounds believe that they do not work

and that we should not be doing them (Biesta, 2015; Cartwright, 2019). Researchers and policymakers have strongly held views about RCTs and about the kinds of scientific evidence needed to support warranted claims in education. These views are likely to be shaped by individual paradigm and discipline allegiances and by professional, social, and ethical value systems.

Beyond the philosophical and paradigmatic debates, concerns have been raised about the epistemological appropriateness of using RCTs in education (Biesta, 2007, 2010, 2015; Black, 2018; Dawson et al., 2018; Deaton & Cartwright, 2018; Gorard et al., 2017; Gough, 2021; Greenleaf & Petrosino, 2009; Lortie-Forgues & Inglis, 2019; Ming & Goldenberg, 2021; Rowe & Oltmann, 2016; Thomas, 2016; Wolf & Harbatkin, 2023; Wrigley & McCusker, 2019). We have already discussed the validity of outcome measures and the generalisability of RCT-detected effects across educational settings. Other specific concerns are discussed below and include: the ethical nature of randomisation; the challenge of blinding participants in RCTs; doubts about the fidelity of intervention implementation across different settings; lack of integration with relevant theory; and policymakers' concerns that RCTs are too difficult, too costly, and too slow.

Ethical Dilemma of Randomisation

RCTs have been criticised on ethical grounds because they create an unacceptable dilemma by randomly depriving some students from potentially useful interventions, or randomly exposing some students to potentially harmful or worthless interventions (Cartwright, 2011; Cook, 2003; Dawson et al., 2018). In response, one could ask how the potential moral harm resulting from randomisation compares to the harm resulting from introducing or maintaining unproven interventions in schools that may have little or no benefit. If the effectiveness of interventions, as determined by RCTs, influences decisions that benefit all students in the experiment and potentially a much wider population, how then would we assess the ethical dilemma?

Blinding of Trial Participants

To prevent participants' behaviour being influenced by the knowledge that they are taking part in a trial, it is good practice that blinding be done in RCTs. There is a real challenge with blinding participants to their involvement in RCTs in an educational setting, where teachers and students in experimental groups are likely to know they are receiving an intervention (Hutchison & Styles, 2010; Styles & Torgerson, 2018). In many RCTs, teachers have a role in delivering an intervention of interest, adding further to the blinding challenge. It is more practical for investigators in an RCT to be blinded as to who has received an intervention. This is particularly important so that members of a research team involved in collecting outcome data avoid unconscious information bias.

Fidelity of Intervention Implementation

Given the nature of education as a complex social process, there are doubts about the fidelity of RCT implementation (Clyne et al., 2020). Accuracy and transparency in the design, conduct, and reporting of RCTs are essential to their reputation as a source of evidence for decision-making (Edovald & Nevill, 2020; Hutchison & Styles, 2010; Sims et al., 2023). Publishing the protocols used and the quality standards for the implementation of individual RCTs should be a minimum requirement. Compliance with international standards, such as the CONSORT[1] statement and the PRISMA[2] guidelines, can help to promote confidence in findings and facilitates systematic reviews and meta-analysis of existing RCTs. For educational researchers, the WWC in the US has published a procedures and standards handbook for conducting RCTs and a set of reporting guidelines for study authors (WWC, 2022). While deficiencies remain, these and other efforts to improve the fidelity of intervention implementation in RCTs can improve reproducibility and reduce waste in research funding.

Atheoretical Status of RCTs

Researchers are rightly criticised if they do not attempt to integrate existing theoretical knowledge into their RCT design or do not use RCTs to contribute to the development and testing of theory (Connolly et al., 2018). When RCTs are combined with theoretical knowledge, they are more likely to produce better explanations and better evaluations, and where RCTs falsify causal assumptions, they can contribute to the development, testing, and refinement of educational theory (Deaton & Cartwright, 2018; Sims et al., 2023). RCTs supported by the EEF are designed to evaluate both the educational theory underlying an approach, and a specific intervention or programme (Dawson et al., 2018). The logic of this dual mandate is to develop a theoretically informed idea into a practical initiative that can be used in an educational setting. Process evaluations, which examine how an intervention of interest was implemented under experimental conditions, should be conducted as part of any RCT. Process evaluations typically use qualitative research such as interviews to explain the causal processes underlying RCT results and can therefore add to our theoretical understanding about *how* a specific intervention causes change. Subgroup analysis examines the effectiveness of interventions on target cohorts of students and can provide more discreet and theoretical accounts about underlying causal assumptions. Hence, integrating theoretical knowledge into the design of RCTs and using process evaluations and subgroup analysis can contribute to theory development and produce richer explanatory models. Integrating RCTs with available theory means it is possible to move beyond narrow questions about "what works" and ask how an intervention worked, for whom it worked, and under what conditions (Connolly et al., 2018).

1 CONSORT Statement – Guidelines on reporting RCTs https://www.equator-network.org/reporting-guidelines/consort/
2 PRISMA Guidelines on reporting meta-analysis http://www.prisma-statement.org/

Costs and Timing of RCTs

The cost of undertaking RCTs and meta-analysis in education is relatively high. The expertise and skills required to undertake RCTs take time to acquire and RCTs can take years to complete, making published findings less relevant for policymakers. The cost of RCTs reduces as practice increases and there is some evidence that recent advances in artificial intelligence and the nascent technology of large language models have the potential to further reduce the cost of synthesis studies (Tyler et al., 2023). The commissioning and synthesis of RCTs are a significant investment by each country's government which needs to be considered in terms of long-term ambition for using trustworthy and impartial knowledge about the effectiveness of education policies and classroom practice.

Based on the foregoing discussion, there are cases where a rigorously conducted RCT, even when integrated with other forms of educational research, will not be the most appropriate method to evaluate an education policy or practice. For example, government attempts to address educational disadvantage involve many different policies and programmes all mediated through a complex web of interactions among students, teachers, schools, and families. This plays out within a wider set of social and economic interests and influences (Brown & Malin, 2022; Gilleece & Clerkin, 2024; Gleeson et al., 2023). For "wicked" policy problems, RCTs may be impractical, and observational studies may be more appropriate, even if they do not demonstrate causality (Hutchison & Styles, 2010; Rawdon et al., 2020; Styles & Torgerson, 2018).

RCTs in Education in Ireland

In Ireland, RCTs are commonly used to inform many areas of public policy and associated practice. Practice is most apparent in healthcare, where RCTs are the bedrock of clinical research (see for example the *Clinical Trials Network* [https://ncto.ie/clinical-trial-networks/], and *Evidence Synthesis Ireland* [https://evidencesynthesisireland.ie/], both funded by the Health Research Board). RCTs have also been used in Ireland to test and evaluate policy options in economics; implement reforms in social protection; and better understand and influence our behaviour as energy consumers (Byrne et al., 2022; Lunn & Robertson, 2018; Purcell, 2016; Timmons & Lunn, 2023). In 2023, the Department for Children, Equality, Disability, Integration and Youth established the *What Works Ireland Evidence Hub* (https://whatworks.gov.ie/), a repository of published evidence reviews on the effectiveness of prevention and early intervention programmes aimed at improving outcomes for children. Evaluations are based on results published by the Early Intervention Foundation in the UK, where only RCTs and quasi-experimental studies comply with the highest standard for evidence.

Ireland does not have a strong culture of systematic evaluation in education and has no tradition of conducting large-scale experimental or quasi-experimental educational

studies (Gilleece & Clerkin, 2024). There are some exceptions, such as the evaluation of the FRIENDS for life emotional resilience programme in Irish primary schools (Ruttledge et al., 2016), and the evaluation of a literacy after-school programme (Biggart et al., 2012). RCTs on the school-going population in Ireland are more likely to have been undertaken by health researchers interested in aspects of physical or mental health. See, for example, Martin and Murtagh's (2015) "Active classrooms", a cluster RCT, where classes rather than individual students were the observable units facilitating the measurement of the community effect of integrating physical activity into classroom lessons, or Staines et al.'s (2023) synthesis of RCT findings on school-based interventions aimed at reducing suicidal behaviour.

The reasons for the low incidence of RCTs in education in Ireland are unclear. A recent publication by Gilleece and Clerkin (2024) on the prospects for using RCTs in evaluating the DEIS[3] programme is very welcome and fitting too that the authors are affiliated with the Educational Research Centre, an institution well-placed to take a deep dive into RCTs in education. Any critical debate on the topic would be well overdue not least since the paradigm wars during the 1970s and 1980s were a muted affair in Ireland, played out against the backdrop of the international turn against experimental studies in education. As Sugrue (2009) points out, the international "shakedown between the empiricist instrumentalists and the liberal progressives" (p. 9) was a non-event in Ireland with the progressives having things mostly their own way. Ó Súilleabháin (1981) captured the prevailing anti-positivist feeling in Ireland at the time concluding that "educational research is of an order quite different from that of the experimental or mathematical sciences" (p. 7). In a subsequent typology of educational studies in Ireland, Kellaghan (1989) reported a dearth of experimental or quasi-experimental studies capable of providing an "estimation of the relationships between variables in an attempt to increase understanding of the phenomena under investigation" (p. 198). Ireland has seen little critical commentary in response to the "what works" movement in education, the international rise of RCTs, and hegemonic visions of educational research more generally. Elsewhere, apathy towards RCTs and EBP has prompted questions about the rigour and relevance of educational research (Cartwright, 2019; Cook, 2002, 2003; Hargreaves, 1996; Hillage et al., 1998; Lather, 2008; Nutley et al., 2013; Oancea, 2005; O'Connor, 2018; Perry et al., 2022; Tooley & Darby, 1998; Susen, 2015; Van Damme, 2022; Welsh, 2021).

In one of the few insider accounts about the use of educational research by policymakers in Ireland, the author collected data from a series of semi-structured interviews with senior government officials in education and with educational researchers in Irish universities and the two main research institutes, namely the Educational Research Centre and the Economic and Social Research Institute (O'Connor, 2018). A snapshot of these testimonies suggests a general scepticism about the value of RCTs in education (see Box 1).

3 Delivering Equality of Opportunity in Schools (DEIS) is a government programme designed to give tailored support to schools which have high concentration of disadvantage.

BOX 1

Educational Researchers' and Policymakers' Views on RCTs in Ireland

Views of Educational Researchers

The only place they [RCTs] have got traction here is [in] the Atlantic Philanthropy funded projects. I think there would be significant resistance on ethical grounds for [this approach] being used more widely in Ireland (Research Institute Researcher #2)

I would be kind of suspicious. I think too much is placed on RCTs, they have become very popular internationally, the clearing house idea and the idea that if you do x then y will follow. I think [this approach] sometimes tries to offer instant solutions for things that are not as tractable as it would appear (Research Institute Researcher #2)

Our evaluation of DEIS, reports prepared in particular over the last five years, have been roundly criticised by a small number of commentators on the ground that it is flawed because there is no control group (Research Institute Researcher #4)

We are faced with a situation where we have a pot of money for disadvantage and what are we going to do? Are we going to identify a group of schools, withhold treatment from half of them and then give it to the other half? (Research Institute Researcher #4)

We don't do them [RCTs] because they don't work; you have to come at [this] in another way that may not be as conclusive as how they do things in medicine but that's what we have to work with (University Researcher #4)

It is usually ethically and methodologically where it sort of falls apart for us. If you do it [an RCT] in a naturally occurring setting you have a methodological mess because you can't control so much…and ethically you shouldn't really be putting kids in different conditions because [this can] affect them…Schools are messy complex places. You can't really control those environments to 95% certainty. The best you can say is that there is a bit of an association between that and that, that's not what policymakers want to hear (University Researcher #7)

Views of Government Officials in Department of Education

I suppose there is the challenge always in education and in educational research in terms of say the control group versus the actual experimental group… [if you think of] the likes of say Ben Goldacre and the stuff that he has written in Britain about the rigours of scientific research and how it applies to the social sciences (Government Official #1)

I knew that there are influential decision-makers in powerful Government Departments who really like them [RCTs] and think that educators are dodging them (Government Official #3)

I get purveyors sitting at my table saying that they have proven that some practice works and why won't teachers do it? I say that teachers don't like it; it doesn't matter that it works, you didn't talk to them, you used some test (Government Official #5)

Conclusion

In Ireland, government enthusiasm for evidence-informed policy and practice has led to new structures and mechanisms across government departments aimed at strengthening the alignment between research, policy, and practice (OECD, 2023b). The current political backdrop is supportive of policymakers and school leaders in education who are committed to putting scientific evidence at the heart of policy and practice in education to achieve better outcomes for students and make decision-making processes more transparent and democratically accountable. International interest in "what works" in education has led many governments to commission RCTs as a method aimed at providing impartial evaluations of the effectiveness of policy decisions and classroom-based practice. Decision-makers require holistic accounts and explanations about effective approaches to teaching and learning or effective programmes in education that improve equity and social inclusion outcomes. Research methods appropriate to inquiry in the natural sciences are not automatically well suited to the study of education, where the learning context is much harder to predict or control and where policymakers and practitioners need to know not just what works but also what matters and what is acceptable.

The use of RCTs in education is controversial and their status is contested, even among those who otherwise agree that policy and practice should be informed by available research. Perspectives on RCTs are quite polarised and their contribution in education is often assessed through a rigid dichotomy of good or bad science, but this creates flawed, simplistic, and caricatured narratives of what counts as scientific evidence (Ravitch, 2005; Wrigley & McCusker, 2019). A binary representation and portrayal of positivist and interpretivist forms of research in education can force us to take up positions about the legitimacy or otherwise of applying natural science research methods to the systematic study of the social world. A consequence of polarised debates is that RCTs are either too readily oversold or too easily discredited. Education is a complex social process where problems are multidimensional and untangling them requires integrated and holistic understandings informed by mixing methods of inquiry and drawing upon multidisciplinary sources of knowledge. Users of educational research should be pragmatic and non-partisan when it comes to selecting research approaches most appropriate to addressing identified knowledge gaps in policy and practice.

Myths, misunderstandings, and false claims about RCTs should not go unchallenged and critical engagement should steer well clear of lazy hierarchies or rankings of educational research, and instead see RCTs for what they are capable of and what role they can have within the full spectrum of educational research. There are many pressing reasons to systematically map and assess the national capacity for generating, mobilising, and using educational research to improve policy and practice in Ireland. Examining the role of RCTs in education is an important part of this wider reflection on

the contribution of educational research. The outcomes of these kinds of discussions may be inconvenient for the administration and for the research community, so the question is: Do we really want to know?

References

Armytage, H. (1976). Educational research: A tercentenary romp. *Research Intelligence*, *2*(2), 5–13. https://doi.org/10.1080/0141192760020202

Baird, M. D., & Pane, J. F. (2019). Translating standardized effects of education programs into more interpretable metrics. *Educational Researcher*, *48*(4), 217–228. https://doi.org/10.3102/0013189X19848729

Ball, S. J. (1997). Policy sociology and critical social research: A personal review of recent educational policy and policy research. *British Educational Research Journal*, *23*(3), 257–274. https://www.jstor.org/stable/1502044

Ball, S. J. (2001). 'You've been NERFed!' Dumbing down the academy: National Educational Research Forum: 'A national strategy – consultation paper': A brief and bilious response. *Journal of Education Policy*, *16*(3), 265–268. http://dx.doi.org/10.1080/02680930110041393

Ball, S. J. (2015). Education, governance and the tyranny of numbers. *Journal of Education Policy, 30*(3), 299–301. https://doi.org/10.1080/02680939.2015.1013271

Bennett, T. (2013). *Teacher proof: Why research in education doesn't always mean what it claims, and what you can do about it*. Routledge.

Bergeron, P.-J., & Rivard, L. (2017). How to engage in pseudoscience with real data: A criticism of John Hattie's arguments in *Visible Learning* from the perspective of a statistician. *McGill Journal of Education*, *52*(1), 237–246. https://doi.org/10.7202/1040816ar

Biesta, G. J. J. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, *57*(1), 1–22. https://doi.org/10.1111/j.1741-5446.2006.00241.x

Biesta, G. J. J. (2010). Why 'what works' still won't work: From evidence-based education to value-based education. *Studies in Philosophy and Education*, *29*, 491–503. https://doi.org/10.1007/s11217-010-9191-x

Biesta, G. J. J. (2015). Improving education through research? From effectiveness, causality and technology to purpose, complexity and culture. *Policy Futures in Education*, *14*(2), 194–210. https://doi.org/10.1177/1478210315613900

Biggart, A., Kerr, K., O'Hare, L., & Connolly, P. (2012). *Evaluation of the effectiveness of the childhood development initiative's Doodle Den literacy programme*. Childhood Development Initiative. http://hdl.handle.net/10147/299260

Black, C. (2018). *Science/fiction: 'Evidence-based' education, scientific racism, & how learning styles became a myth*. http://carolblack.org/science-fiction/

Blass, E. (2020). Debunking Hattie: Evaluating the contribution of academic studies to policy development and implementation in Australia. *Journal of Education & Social Policy*, *7*(4), 91–97. https://doi.org/10.30845/jesp.v7n4p8

Brown, C., & Malin, J. R. (Eds.). (2022). *The emerald handbook of evidence-informed practice in education: Learning from international contexts*. Emerald Publishing Limited.

Byrne, S., Devine, K., & McCarthy, Y. (2022). *Behavioural economics and public policy-making*. Quarterly Bulletin 01/January 2022, Central Bank of Ireland. https://www.centralbank.ie/docs/default-source/publications/quarterly-bulletins/quarterly-bulletin-signed-articles/behavioural-economics-and-public-policy-making.pdf?sfvrsn=7

Cairney, P. (2016). *The politics of evidence-based policy making*. Palgrave Pivot.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally & Company.

Cartwright, N. (2011). A philosopher's view of the long road from RCTs to effectiveness. *The Lancet, 377*(9775), 1400–1401. https://doi.org/10.1016/s0140-6736(11)60563-1

Cartwright, N. (2019). What is meant by "rigour" in evidence-based educational policy and what's so good about it? *Educational Research and Evaluation*, *25*(1–2), 63–80. https://doi.org/10.1080/13803611.2019.1617990

Clark, C. (2005). The structure of educational research. *British Educational Research Journal*, *31*(3), 289–308. https://doi.org/10.1080/01411920500082128

Clyne, B., Boland, F., Murphy, N., Murphy, E., Moriarty, F., Barry, A., Wallace, E., Devine, T., Smith, S. M., Devane, D., Murphy, A., & Fahey, T. (2020). Quality, scope and reporting standards of randomised controlled trials in Irish Health Research: An observational study. *Trials, 21*, 494. https://doi.org/10.1186/s13063-020-04396-x

Cochran-Smith, M., & Lytle, S. (2006). Troubling images of teaching in No Child Left Behind. *Harvard Educational Review*, *76*(4), 668–697. https://doi.org/10.17763/haer.76.4.56v8881368215714

Coe, R. (2002). *It's the effect size, stupid: What effect size is and why it is important*. Proceedings of the British Educational Research Association Annual Conference. https://f.hubspotusercontent30.net/hubfs/5191137/attachments/ebe/ESguide.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Lawrence Erlbaum Associates.

Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980–2016. *Educational Research, 60*(3), 276–291. https://doi.org/10.1080/00131881.2018.1493353

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered and for not doing them. *Educational Evaluation and Policy Analysis*, *24*(3), 175–199. https://www.jstor.org/stable/3594164

Cook, T. D. (2003). Why have educational evaluators chosen not to do randomized experiments? *The ANNALS of the American Academy of Political and Social Science*, *589*(1), 114–149. https://doi.org/10.1177/0002716203254764

Cope, E. (1975). Practical experience in the professional training of teachers: An overview of some research issues. *Research Intelligence*, *1*(2), 71–72. https://www.jstor.org/stable/1520828

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications. Inc.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*(2), 116–127. https://doi.org/10.1037/h0076829

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. SAGE Publications Ltd.

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97–140. https://doi.org/10.1080/10888691.2018.1537791

Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies, 47*(2), 108–121. https://doi.org/10.1111/1467-8527.00106

Dawson, A., Yeomans, E., & Brown, E. R. (2018). Methodological challenges in education RCTs: Reflections from England's Education Endowment Foundation. *Educational Research, 60*(3), 292–310. https://doi.org/10.1080/00131881.2018.1500079

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine, 210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Edovald, T., & Nevill, C. (2020). Working out what works: The case of the Education Endowment Foundation in England. *ECNU Review of Education, 4*(1), 46–64. https://doi.org/10.1177/2096531120913039

European Commission. (2007). *Commission staff working document: Towards more knowledge-based policy and practice in education and training*. SEC(2007) 1098. https://op.europa.eu/en/publication-detail/-/publication/962e3b89-c546-4680-ac84-777f8f10c590

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher, 31*(8), 4–14. https://www.jstor.org/stable/3594387

Gilleece, L., & Clerkin, A. (2024). Towards more robust evaluation of policies and programmes in education: Identifying challenges in evaluating DEIS and Reading Recovery. *Irish Educational Studies*. https://doi.org/10.1080/03323315.2024.2334704

Gleeson, J. (2012). The professional knowledge base and practice of Irish post-primary teachers: What is the research evidence telling us? *Irish Educational Studies*, *31*(1), 1–17. https://doi.org/10.1080/03323315.2011.617945

Gleeson, J., Rickinson, M., Walsh, L., Cutler, B., Salisbury, M., Hall, G., & Khong, H. (2023). Quality use of research evidence: Practitioner perspectives. *Evidence & Policy*, *19*(3), 423–443. https://doi.org/10.1332/174426421X16778434724277

Goldacre, B. (2013). *Building evidence into education*. Department for Education (UK). https://www.gov.uk/government/news/building-evidence-into-education

Gorard, S., See, B. H., & Siddiqui, N. (2017). *The trials of evidence-based education: The promises, opportunities and problems of trials in education*. Routledge.

Gough, D. (2021). Appraising evidence claims. *Review of Research in Education*, *45*(1), 1–26. https://doi.org/10.3102/0091732X20985072

Greenleaf, C., & Petrosino, A. (2009). Response to Slavin, Cheung, Groff, and Lake (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly*, *44*(4), 349–354. https://doi.org/10.1598/RRQ.44.4.2

Hammersley, M. (2003). Can and should educational research be educative? *Oxford Review of Education, 29*(1), 3–25. https://www.jstor.org/stable/1050726

Hargreaves, A. (1996). Transforming knowledge: Blurring the boundaries between research, policy, and practice. *Educational Evaluation and Policy Analysis*, *18*(2), 105–122. https://doi.org/10.3102/01623737018002105

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J. (2023). *Visible learning: The sequel. A synthesis of over 2,100 meta-analyses relating to achievement*. Routledge.

Haynes, L., Service, O., Goldacre, B., & Torgerson, D. J. (2012). *Test, learn, adapt: Developing public policy with randomised controlled trials*. Cabinet Office - Behavioural Insights Team. http://dx.doi.org/10.2139/ssrn.2131581

Hedges, L. V., & Schauer, J. (2018). Randomised trials in education in the USA. *Educational Research, 60*(3), 265–275. https://doi.org/10.1080/00131881.2018.1493350

Hillage, J., Pearson, R., Anderson, A., & Tamkin, P. (1998). *Excellence in research on schools*. DfEE Research Report 74. Department for Education and Employment (UK). https://dera.ioe.ac.uk/id/eprint/9856

Hudson, G. (1975). Two paradigms of educational research? *Research Intelligence*, *1*(2), 68–71. https://www.jstor.org/stable/1520827

Hutchison, D., & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers*. National Foundation for Educational Research. https://www.nfer.ac.uk/publications/a-guide-to-running-randomised-controlled-trials-for-educational-researchers/

Kellaghan, T. (1989). The interface of research, evaluation and policy in Irish education. In D. G. Mulcahy & D. O'Sullivan (Eds.), *Irish educational policy: Process and substance* (pp. 191–218). Institute of Public Administration.

Kicinski, M. (2014). How does under-reporting of negative and inconclusive results affect the false-positive rate in meta-analysis? A simulation study. *BMJ Open, 4*, e004831. https://doi.org/10.1136/bmjopen-2014-004831

Lather, P. (2008). New wave utilization research: (Re)imaging the research/policy nexus. *Educational Researcher, 37*(6), 361–364. https://www.jstor.org/stable/25209012

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, *48*(3), 158–166. https://doi.org/10.3102/0013189X19832850

Lunn, P. D., & Robertson, D. (2018). *Using behavioural experiments to pre-test policy*. Budget Perspectives 2019, Paper 2. Economic & Social Research Institute. https://doi.org/10.26504/bp201902

Martin, R., & Murtagh, E. (2015). An intervention to improve the physical activity levels of children: Design and rationale of the 'Active Classrooms' cluster randomised controlled trial. *Contemporary Clinical Trials*, *41*, 180–191. https://doi.org/10.1016/j.cct.2015.01.019

Ming, N. C., & Goldenberg, L. B. (2021). Research worth using: (Re)framing research evidence quality for educational policymaking and practice. *Review of Research in Education, 45*(1), 129–169. https://doi.org/10.3102/0091732X21990620

Nutley, S., Powell, A., & Davies, H. (2013). *What counts as good evidence? Provocation paper for the alliance for useful evidence*. Alliance for Useful Evidence. https://www.nesta.org.uk/report/what-counts-good-evidence/

Oancea, A. (2005). Criticisms of educational research: Key topics and levels of analysis. *British Educational Research Journal*, *31*(2), 157–183. http://dx.doi.org/10.1080/01411920052000340198

OECD. (2007). *Evidence in education: Linking research and policy*. OECD Publishing. https://doi.org/10.1787/9789264033672-en

OECD. (2022). *Who cares about using education research in policy and practice? Strengthening research engagement*. Educational Research and Innovation. OECD Publishing. https://doi.org/10.1787/d7ff793d-en

OECD. (2023a). *Shaping digital education: Enabling factors for quality, equity and efficiency*. OECD Publishing. https://doi.org/10.1787/bac4dc9f-en

OECD. (2023b). *Strengthening policy development in the public sector in Ireland*. OECD Public Governance Reviews. OECD Publishing. https://doi.org/10.1787/6724d155-en

OECD. (2023c). *Who really cares about using education research in policy and practice? Developing a culture of research engagement*. Educational Research and Innovation. OECD Publishing. https://doi.org/10.1787/bc641427-en

Oliver, S., Bagnall, A. M., Thomas, J., Shepherd, J., Sowden, A., White, I., Dinnes, J., Rees, R., Colquitt, J., Oliver, K., & Garrett, Z. (2010). Randomised controlled trials for policy interventions: A review of reviews and meta-regression. *Health Technology Assessment, 14*(16), 1–165. https://doi.org/10.3310/hta14160

Ozga, J. (2012). Governing knowledge: Data, inspection and education policy in Europe. *Globalisation, Societies and Education*, *10*(4), 439–455. https://doi.org/10.1080/14767724.2012.735148

O'Connor, J. (2018). *Education policymaking in Ireland: Does research matter?* [Unpublished doctoral thesis]. University College Dublin.

O'Connor, J. (2024). Evidence based education policy in Ireland: Insights from educational researchers. *Irish Educational Studies*, *43*(1), 21–45. https://doi.org/10.1080/03323315.2021.2021101

Ó Súilleabháin, S. V. (1981). What are educational studies? *Irish Educational Studies*, *1*(1), 1–8. https://doi.org/10.1080/0332331810010105

Pampaka, M., Williams, J., & Homer, M. (2016). Is the educational 'what works' agenda working? Critical methodological developments. *International Journal of Research & Method in Education, 39*(3), 231–236. https://doi.org/10.1080/1743727X.2016.1170476

Pegram, J., Watkins, R. C., Hoerger, M., & Hughes, J. C. (2022). Assessing the range and evidence-base of interventions in a cluster of schools. *Review of Education*, *10*(1), e3336. https://doi.org/10.1002/rev3.3336

Pellegrini, M., & Vivanet, G. (2021). Evidence-based policies in education: Initiatives and challenges in Europe. *ECNU Review of Education, 4*(1), 25–45. https://doi.org/10.1177/2096531120924670

Perry, T., Morris, R., & Lea, R. (2022). A decade of replication study in education? A mapping review (2011–2020). *Educational Research and Evaluation*, *27*(1–2), 12–34. https://doi.org/10.1080/13803611.2021.2022315

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Rowman & Littlefield.

Purcell, K. (2016). *Staff paper 2016: Applying behavioural economics in Irish policy*. Department of Public Expenditure and Reform. https://assets.gov.ie/181486/bb1f83ae-3ea4-4f60-a196-99fd515eb73a.pdf

Ravitch, D. (Ed.). (2005). *Brookings papers on education policy*. Brookings Institution Press.

Rawdon, C., Sampson, K., Gilleece, L., & Cosgrove, J. (2020). *Developing an evaluation framework for teachers' professional learning in Ireland: Phase 1 desk-based research*. Educational Research Centre. https://www.erc.ie/wp-content/uploads/2020/11/ERC-TPL-Wellbeing-Literature-Review-Report_Final.pdf

Ropovik, I., Adamkovic, M., & Greger D. (2021). Neglect of publication bias compromises meta-analyses of educational research. *PLoS One*, *16*(6), e0252415. https://doi.org/10.1371/journal.pone.0252415

Rowe, M., & Oltmann, C. (2016). Randomised controlled trials in educational research: Ontological and epistemological limitations. *African Journal of Health Professions Education, 8*(1), 6–8. http://dx.doi.org/10.7196%2FAJHPE.2016.v8i1.683

Ruttledge, R., Devitt, E., Greene, G., Mullany, M., Charles, E., Frehill, J., & Moriarty, M. (2016). A randomised controlled trial of the FRIENDS for Life emotional resilience programme delivered by teachers in Irish primary schools. *Educational & Child Psychology*, *33*(2), 69–89. https://doi.org/10.53841/bpsecp.2016.33.2.69

Sims, S., Anders, J., Inglis, M., Lortie-Forgues, H., Styles, B., & Weidmann, B. (2023). *Experimental education research: Rethinking why, how and when to use random assignment* (CEPEO Working Paper No. 23–07). Centre for Education Policy and Equalising Opportunities, University College London. https://EconPapers.repec.org/RePEc:ucl:cepeow:23-07

Siraj-Blatchford, I., & Siraj-Blatchford, J. (1997). Reflexivity, social justice and educational research. *Cambridge Journal of Education*, *27*(2), 235–248. https://doi.org/10.1080/0305764970270207

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, *31*(7), 15–21. https://www.jstor.org/stable/3594400

Slavin, R. E. (2008a). Evidence-based reform in education: What will it take? *European Education Research Journal, 7*(1), 124–128. https://doi.org/10.2304/eerj.2008.7.1.124

Slavin, R. E. (2008b). Perspectives on evidence-based research in education – What works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5–14. https://doi.org/10.3102/0013189X08314117

Slavin, R. E. (2020). How evidence-based reform will transform research and practice in education. *Educational Psychologist*, *55*(1), 21–31. https://doi.org/10.1080/00461520.2019.1611432

Spybrook, J., Zhang, Q., Kelcey, B., & Dong, N. (2020). Learning from cluster randomized trials in education: An assessment of the capacity of studies to determine *what works*, *for whom*, and *under what conditions*. *Educational Evaluation and Policy Analysis*, *42*(3), 354–374. https://doi.org/10.3102/0162373720929018

Staines, L., Healy, C., Corcoran, P., Keeley, H., Coughlan, H., McMahon, E., Cotter, P., Cotter, D., Kelleher, I., Wasserman, C., Brunner, R., Kaess, M., Sarchiapone, M., Hoven, C. W., Carli, V., Wasserman, D., & Cannon, M. (2023). Investigating the effectiveness of three school-based interventions for preventing psychotic experiences over a year period – a secondary data analysis study of a randomized controlled trial. *BMC Public Health, 23*, 219. https://doi.org/10.1186/s12889-023-15107-x

Styles, B., & Torgerson, C. J. (2018). Randomised controlled trials (RCTs) in education research – *methodological debates, questions, challenges*. *Educational Research*, *60*(3), 255–264. https://doi.org/10.1080/00131881.2018.1500194

Sugrue, C. (2009). Back to the future: Perspectives on current realities and future quality of educational research through the prism of *Irish Educational Studies*. *Irish Educational Studies*, *28*(1), 7–25. https://doi.org/10.1080/03323310802597283

Susen, S. (2015). *The 'postmodern turn' in the social sciences.* Palgrave Macmillan. https://doi.org/10.1057/9781137318237

Terhart, E. (2011). Has John Hattie really found the holy grail of research on teaching? An extended review of *Visible Learning*. *Journal of Curriculum Studies, 43*(3), 425–438. https://doi.org/10.1080/00220272.2011.576774

Thomas, G. (2016). After the gold rush: Questioning the "gold standard" and reappraising the status of experiment and randomized controlled trials in education. *Harvard Educational Review*, *86*(3), 390–411. https://doi.org/10.17763/1943-5045-86.3.390

Timmons, S., & Lunn, P. D. (2023). Behaviourally-informed household communications increase uptake of radon tests in a randomised controlled trial. *Scientific Reports, 13*, 20401. https://doi.org/10.1038/s41598-023-47832-z

Tooley, J., & Darby, D. (1998). *Educational research: A critique – A survey of published educational research*. Office for Standards in Education. http://hdl.voced.edu.au/10707/137550

Torgerson, D. J., & Torgerson, C. J. (2008). *Designing randomised trials in health, education and the social sciences: An introduction*. Palgrave Macmillan.

Torgerson, C. J. & Torgerson, D. J. (2013). *Randomised controlled trials in education: An introductory handbook.* Educational Endowment Foundation.

Tyler, C., Akerlof, K. L., Allegra, A., Arnold, Z., Canino, H., Doornenbal, M. A., Goldstein, J. A., Pedersen, D. B., & Sutherland, W. J. (2023). AI tools as science policy advisers? The potential and the pitfalls. *Nature*, *622*, 27–30. https://doi.org/10.1038/d41586-023-02999-3

UNESCO. (2023). *Global education monitoring report* 2023 – *Technology in education: A tool on whose terms?* UNESCO. https://www.unesco.org/gem-report/en/technology

Van Damme, D. (2022). *The power of proofs: (Much) beyond RCTs*. Centre for Curriculum Redesign. https://curriculumredesign.org/wp-content/uploads/The-Power-of-Proofs.pdf

Welsh, R. O. (2021). Assessing the quality of education research through its relevance to practice: An integrative review of research-practice partnerships. *Review of Research in Education*, *45*(1), 170–194. https://doi.org/10.3102/0091732X20985082

What Works Clearinghouse. (2022). *What Works Clearinghouse: Procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Handbooks

White, H. (2019). The twenty-first century experimenting society: The four waves of the evidence revolution. *Palgrave Communications, 5*, 47. https://doi.org/10.1057/s41599-019-0253-6

Whitty, G. (2006). Education(al) research and education policy making: Is conflict inevitable? *British Educational Research Journal, 32*(2), 159–176. https://doi.org/10.1080/01411920600568919

Wolf, B., & Harbatkin, E. (2023). Making sense of effect sizes: Systematic differences in intervention effect sizes by outcome measure type. *Journal of Research on Educational Effectiveness*, *16*(1), 134–161. https://doi.org/10.1080/19345747.2022.2071364

Wrigley, T., & McCusker, S. (2019). Evidence-based teaching: A simple view of "science". *Educational Research and Evaluation*, *25*(1-2), 110–126. https://doi.org/10.1080/13803611.2019.1617992