

**Crook:** Well, I'm really pleased to introduce Gary Thomas from the University of Birmingham. We're going to talk about his recent paper. Details of that paper will be attached to the page, which its recording is on. The paper is questioning the contribution of experiments and experimental method to educational inquiry. We can't do it for justice in a 30 minute conversation, but I hope this helps others decide to read it. They have access to it or and/or to talk about it elsewhere themselves. So let's get going. Maybe I could start with a little bit of semantic groundwork, Gary. The paper starts with some head scratching about the meaning of the term experiment. There is a kind of, I think, an approving nod to a child's understanding. That experiment might mean an exercise in finding out. So I think the reader feels that the term experiment can still have some useful currency, and the paper is perhaps simply questioning one version of its meaning as a research practice within education. And so questioning the utility of conclusions that can be generated by education experiments now, is that a fair starting point?

**Thomas:** Oh, yes, I think so. You've you said to me that you think that we can perhaps distinguish between experiment with a capital E and with a little E. I think that very nicely sums up what the article is about the fact that the idea of experiment has been appropriated by a particular interest group. In such a way that we think of it, not in the way that natural scientists think of experiment. I mean, that's the way I start off really in the article talking about the fluidity, the flexibility, the relaxation with which scientists, natural scientists talk about experiment and the difference therein. Between that and the way that educators and social scientists talk about experiment seems to me that we've kind of become corralled in this particular view of experiment, which is that you you compare groups, one of which you do something to the other of which you don't do anything to, and you see if there's any difference between the two. Now that seems to me to be a perfectly acceptable way of doing experiment if you're in plant science or pharmaceuticals. I mean, I wouldn't want to take any drug that hadn't been tested properly in that way. But, you know, scientists do all kinds of different work using evidence from different fields in different ways to come to conclusions, to develop theory. It seems to me to be, you know, that's not a useful way of going about experiment in education. We should think about it much more flexibly and much more fluidly in education and in and in psychology. I think, you know, we've, as I say, the idea of experiment has become misappropriated, I think it's been taken from a very thin sliver of scientific inquiry, kind of scientific inquiry and and you sort of made it seem as if it's the the best way of doing experimental inquiry and in education.

**Crook:** Okay, that's very helpful. I mean, I remember saying in correspondence with you, the capital in the lower case E - I'm glad that works. In a conversation, of course, we can't really reproduce that. I'm going to suggest that for this conversation, I might revert to the phrase 'formal experiments' to mean the kind of practice that you've just described while allowing the word 'experiment' to embrace a much bigger field. So I mean, even this conversation could be called an experiment. So I'm trying to acknowledge that there's a more generous sense of the term and than something we need to talk about more, which is that, OK, we'll call that a 'formal experiment'.

**Thomas:** Yeah, exactly.

**Crook:** Okay. So one of the things I quite liked in the paper work for me was the idea of you introduce this marketing jargon USP - unique selling point. So maybe it's asking you to put what you just said a slightly different way. But can you share with us how you see the USP of experimental method in education?

**Thomas:** Hmm. Well, I suppose the the USP of experiment is that it establishes clearly a link between cause and effect. I mean, that's it's that's its putative USP in all fields and clearly in some in some fields that's appropriate. In pharmaceuticals that it's appropriate to say that that USP is a valid one. It's its unique selling point. But. Well, the experiment has to be done in particular ways, as there has to be adequate control. Control has to be established properly, which is where randomization comes in so that you've got assignment to groups without any without any bias. There has to be appropriate blinding. For example. But it seems to me that that USP isn't a valid one when we talk about experiment with a capital E in in education. You know, you're not able to to do an experiment in that way in education, it seems to me, I mean, I suppose this is a little bit away from the kind of general thesis that I'm making, but I do make the point that educational experiments don't live up to expectations in this respect. For example, I mean, the term asked randomized controlled trial. It's come to mean not just the randomization of participants between groups, but it's come to mean it's come to embody a whole range of other things that one would expect from an experiment such as adequate double blinding or triple blinding. And when you look at experiments or acts in education, they don't. They don't match up to that. You can't or very rarely. I don't know of any, am willing to be proved wrong. I don't know of any experiments that do or

could be adequately blind the. So, you know, the experimenters themselves and the subjects don't know which groups they're in or triple blind so that the the person doing the evaluation, the person doing the interpretation of the data doesn't know. I suppose that that is perhaps a little bit easier. But in terms of the experimenter and the subject, when the subject is the teacher or the or the children in the classroom. I mean, that's a real problem. I think about the about the randomized controlled trial in education. And you know, there's other issues about size of samples when, when, when medics are talking about RCTs, they're talking about very large groups of up to half a million participants. Those groups seem to me to be much smaller in education and often the participants, the putative participants are in fact grouped by class or some other group. And so it seems to me that it's very difficult to to validly make the draw the conclusions that are drawn from those kinds of experiments.

**Crook:** Yeah, I'd like to come back to just probe those problems a little bit in a moment. I was going to just turn over first, I guess a sense of where you felt this formal experimentation experiments capital E, if you like, was actually flourishing in the education research community. Now you mentioned RCTs. Is that the main space we're considering that is that present themselves as randomized controlled trials, quite large scale projects? Or do you see it flourishing elsewhere?

**Thomas:** Yes. I mean, in a sense, the the term RCT has come to replace experiment, so people tend to think that RCT is something different from experiments. You know, RCTs have become the fashion, not just the fashion, but the principal, the principal kind of funded research by orders of magnitude over the last 10 or 20 years. And yeah, I mean, that's that's really what we're talking about those. I mean, in the United States, there's the 'What Works Foundation' here there's the Education Evaluation Foundation is it? Or what EEF Stands for. But yeah, I mean, that's principally where the funding to kinds of research is going now because.

**Crook:** I guess one one thought that might enter into someone's mind reading the paper is is just how prominent is this mode of research prior to considering its marketing critically? Now, just by way of an example in the issue of British Education Research Journal that carries this paper, there are 14 other research papers, but none of them report a formal experiment. Now, I mean, that can only be suggestive. But but does it

suggest perhaps that formal experimentation is a minority research practice in the discipline?

**Thomas:** It's interesting, yeah. Well, maybe there's a difference between published research and funded research research funded by national governments, by funding bodies and the kinds of research that are published and emerge. Yeah, well, I mean, it's pleasing. I hadn't noticed that that in that particular issue, there are 14 others that aren't about of experimental research. And I don't know what they are particularly about. But yeah, I would like to see more of more action research, more case study research published. But I think. And some of the field I know a little bit about is research and reading and a paper that I've always found fascinating and which challenges the the dominant paradigm, as it's called by the author and paper by Johnston, Peter Johnston, on reading failure. And he uses case study in that in that paper and seems to me to come to some really fascinating conclusions. And he says that the dominant paradigm, you know, the major journals in that field haven't had a single paper that's based on one case study or part of this. No, actually, it wasn't his. His was published in the Harvard Education Review. The major journals have had a, you know, a real prejudice against anything which wasn't scientific. And of course, you know what? What constitutes science is the is the nub of the issue, really. Should science simply be seen as this kind of formally constructed experimental study?

**Crook:** Yeah. I mean, I think an interesting point that's surfacing in my mind from what you've just been saying is what might be a disconnect between something that we encounter as research in the professional journals and books of education and something that is funded by the agencies that supposedly supply resources for research. So does it feel to you to be the case that there's maybe a large community of worker bees, as it were in the education community who are doing the 14 studies that are reported in BERJ, but the front line and the area where research is most likely to be funded is still having a strong expectation that the research should take an eye seat style of approach. Does that seem a reasonable contrast?

**Thomas:** Yeah, I think that sums it up very nicely. And where is that getting us? We're now, we're now getting some evaluations. You, you had an interview yourself with with Matthew Inglis, who's done a very nice evaluation of the evaluations, if you like. It doesn't seem to me it's getting us very far, but that heavy concentration of funding on

research of a particular kind. I suppose that's what's my article about really the where is it getting us? Is it an appropriate form of study? And I said the answer lies in the substrate. The answer lies in the soil. We should look at the kind of soil for the kind of research we want to grow, and that that substrate is very different from the the substrate within which the form of experimental kind of research grew.

**Crook:** Yeah, I like that metaphor. I mean, is it possible, though, to say and I felt that in Matthew English's work and people like Matthew, that there is a potential passage between what you're calling substrate work, maybe the 14 papers that are in that issue of BERJ. Some of them are case studies. Some of them are other forms of .. advertise other forms of method. Is there a natural line between them and the RCT? Is the problem with the style of research that is simply immature and it needs a stronger substrate in place before the necessary work of comparative judgment can be made. Is that a credible analysis?

**Thomas:** Well, I don't know whether I would say that it's immature. You know, I go in in the article to to the history of experimental research in psychology and education over over a hundred years. I talk about the first coming experimentation, the second coming in, the third coming. The first coming was pre-war pre Second World War or between the wars where education experimenters when in a sense, emulating experimenters in psychology. The general consensus, even by proponents at that time, was that the findings were disappointing. They didn't really take us anywhere very meaningful or useful. The second coming was post-war, with people like Campbell and Stanley, who who drew up parameters of different kinds of experiment, came up with unhelpful terms like quasi and true experiment and put. Well, again, the findings were disappointing when very large scale experiments were done principally in the United States and head start following through title one. Which came up with at best inconclusive findings. And Campbell and Stanley Campbell put the inconclusive or not very helpful findings down to lack of randomization. So that second coming was displaced by the third coming and talk about evidence based practice of what works and and introducing randomization. But we're now getting the findings from from those studies that are people like Matthew Inglis have conducted and people like the Leaf and Tainment in the states which are again showing, you know, thoroughly inconclusive results and is that. Talking about the immaturity of the method, I'm not sure it is really the immaturity of the method. I'm not sure whether we can find a better method for doing experimentation. I think the the

essence of it in social situations is just completely inappropriate. I think it was Rossi who talked about the stainless steel law of evaluation, which is the 'better done' paraphrasing, the better done the evaluation, the less likely it is to find anything useful or meaningful. And I think the point I make in the article is that. Is that there is just, you know, the the overwhelming influence of other variables means that the variable of interest is is lost in that sea of other variables?

**Crook:** Okay. Just like to get closure on this before we move on, perhaps to think about other research options. I think I may have been a bit careless in using the term immaturity because I accept and find very useful your kind of historical tracking of the way in which experiment just surfaced in education. That's fine, and it is certainly not immature in that historical sense. But I was wondering if it's if it's intellectually still immature. And the problem might be that the the implications of an asked design are not ready. They've not been made ready by the necessary conceptual work, the substrate where it there needs to be done first. So naive assumptions are being made about what it is that is worth observing. And if theory was richer than an RCT design could be more readily justified now, that would be one position, but it sounds as if you are reluctant to accept the very principle of of a controlled comparison type of research design. Is that right?

**Thomas:** Yes. Yeah, I I find it hard to accept that that one one can do that kind of research, except if you like to disprove it, I suppose it goes back to something deeper, really about what kind of intervention one can make in education or in a social field that is going to have an impact. And I, you know, accept that our experiments are useful in certain circumstances. I was talking with a distinguished user and developer of RCTs in education, Carol, just to hope she doesn't mind me quoting her, Who said that some...She felt that one of the most useful ways in which RCTs can be used is to disprove something to show that something isn't effective. And I think an example that she's written about is if the validation of synthetic phonics in the teaching of reading and the government imposed synthetic phonics on the basis of a few very small studies done in Clackmannanshire, Carol Thordarson not sure whether she did them herself or drew together evidence from randomized controlled trials which showed that synthetic phonics didn't make any difference. Whether the government takes any notice of that or not, or whether it's evidence based practice or practice based evidence is is another issue. But I think it can be useful in showing that that a super duper new technique or

intervention can be useful in showing that it doesn't have an effect, so it has some kind of limited or short term effect.

**Crook:** Yes. I mean, it's interesting contrast because it reminds us that this form of research is not just about what works, it could be about what doesn't work. So here you have a practice already in place, perhaps on rather fragile foundations. And but then aren't you risking appearing to say that we'll accept the outcome of RCTS when they tell us something we're suspicious of as not working, but we don't accept them when they ask us to take something that should work?

**Thomas:** Yeah, well, yes. If they ever do tell you that something should work. And I suppose that's that's not really. I mean, one of the dangers, I think, is that they tell you. They tell you the wrong thing, if something is working, they tell you that it doesn't work. And that's probably partly a reflection on the methodology itself on the way that the trials are actually constructed and conducted. And I think one of the things about the head start research and which led Gene Glass the principle experimenter to say that quantitative experimental evaluations should never be used again was that, you know, these experimental evaluations had shown that head start and follow through and title one didn't work, whereas they probably worked really, really, very well for some people and in some circumstances in those situations. But but the the overall result was that they didn't work. So I suppose they can throw up all sorts of expectations and answers, but. Yeah, I suppose there is the danger of saying it's in the context of that of saying, well, how do we know that that these results were were valid in the in the synthetic phonics studies?

**Crook:** Ok. Let me just I mean, it's a really interesting and feels to me quite a radical claim. So I'm anxious to sort of push the boundaries a little. Now, a moment ago, you use the phrase quasi experiment, I think, or natural experiment. But anyway, I mean, a phrase that captured the idea that there may be conditions in which accidentally or incidentally groups have been created that allow comparisons that look like experiments but have actually arisen by good fortune. I mean, an example that I have in mind you may be able to relate to is Luris's work in Soviet Russia, comparing communities in which schooling had been introduced and communities in which it hadn't been introduced, and then making some quite stark claims about the impact of schooling and its relationship to literacy and so on. Now with that, that made a big difference, I feel, to

the way people thought about literacy and development in the way people thought about the impact of schooling. But would you want to rule it out because it was like an RCT, but a naturally occurring one?

**Thomas:** It's still, you know, I'd go along with Martin Harris's view that inquiry should, should be a matrix of different.... We should be eclectic about the methods that we use and the evidence we collect. You know, it's I think that's been some of the most useful kinds of evidence that we've collected in education in my own field and special education, something similar to to what you've talked about with Luria. You know, states in the United States that have that have had special education states that haven't had special education. And although up to 20 times as much is spent on the pupils in states that have had special education as in those that haven't had special education, equivalent children appear to, you know, have outcomes broadly similar, if not identical. And I think why shouldn't we draw inferences from that? I mean, philosophers of science now talk about science, essentially being about inference to the best explanation, you know, taking in a whole range of different kinds of evidence. And why shouldn't we do that? You know, I talk in the article about. About paleoanthropologists, you know, people who who study the evolution of humankind, who take evidence from from anatomy, from bits of all bone, from fragments of DNA, from Carbon 14 dating and they build narratives which explain or which contribute to a theory which explains and whether it's that or whether it's Darwin, you know, in his finches and his explanation of his theory about evolution or whether it's Einstein and Brownian motion, you know, dropping bits of pollen onto a glass of water. You know, they drew inferences to the best explanation, to the best theory. I don't see why we shouldn't do that. What seems to me to be difficult problematic is when we sort of squeeze the data through particular channels in such a way that it's telling us something confusing and misleading.

**Crook:** Yes. Okay. I think anyone listening would recognize the point about the apparent message of the great figureheads of science, and I'm just thinking who they might be. In education, I mean, one. One name that springs to mind, I suppose, is Piaget. Now, if we're looking for what we might call a more improvisatory disposition in the scientists, maybe a more constructive orientation towards data is Piaget, someone whose work would demonstrate that kind of insight, that kind of synthesis of understanding.



**Thomas:** Yes. Well, I mean, he's he's part of a tradition which, you know, I guess we can draw from. I think, you know, he probably came to some wrong conclusions, but that doesn't mean to say that his general thesis about Constructivism about the child constructing a mental world was wrong. You know, I think his general thesis is one that's contributed to to the way that we think successfully and effectively about about children's thinking. Perhaps one of the problems is that you can come to a faulty conclusions by using. So the improvised methods, but that doesn't mean to say that one shouldn't use improvised methods, it means that other people should should think about this. There's a community of science. There are multiple realities which are essentially boiled down ultimately to a particular theory or a set of theories. So, you know, the idea was critiqued by Vygotsky and Bruner, and the sort of general notion of constructivism is one that survives, whereas the stages that you talked about, so much detail are perhaps best forgotten. But there's, you know, a kind of continuing narrative about the way that the children think that is contributed to very effectively, I think.

**Crook:** Yeah. I mean, one reason Piaget came to my mind, I think, was partly because he seemed to be perhaps a very vivid illustration of the kind of intellect that you might bring to to education or research. But at the same time, as you've just hinted, I think we also now are rather sceptical about some of his particular claims about the stage-like nature of development say, in relation to understanding the physical world and so on. But then I think, well, one of the ways in which we've been led to be skeptical is by a whole load of very ingenious experiments by people like Johnson, Martin Hughes, David Olson and so on. And is that not, in some sense, a vindication of the experimental method?

**Thomas:** Well, again, you know, it's back to the matrix thing, I think, and what we're trying to what we're trying to do. I think if if one's trying to critique something, then we're talking about something rather different experimental method in psychology is different from the experimental method in in education. And I don't think anyone would would argue with the fact that the experiment in in real situations, in real life situations in psychology is it's very different from experiments in the laboratory. And yes, you know, that's contributed to our understanding and critique of of biology methods and his conclusions.

**Crook:** I mean, partly, I'm sort of invoking that kind of example because in my mind is the kind of individual might be listening to this conversation, who is themselves embarking upon a research, education or research trajectory, maybe as a student or maybe as a concerned practitioner? And you ask yourself, well, what kind of methods might they be inclined to embrace? And I think many might be thinking in terms of experiments and what your paper does is, I think, provide a very important caution. So how they might proceed? But but the kind of experiment that those examples are just cited, Olson and Olson and so on is within the reach of a student, I think perhaps doing a Ph.D., but another kind of study that might be in the reach of such a student and whose whose validity I'd be interested in your comment on is actually one of the ones that was in that issue of BERJ, which (you queried, I wonder what they were. I'll tell you about one). It was a report of a multifactor study of teachers who left the profession relating the decision to leave to mental health and wellbeing data that was available for them through the Biobank study. So a large sample of 20000 individuals and so on and led to some conclusions. Basically along the line of the grass isn't actually greener on the other side, (but that doesn't matter). And you probably don't want to comment on a particular study, but it is a species of study and one which is in the reach of a student. I think, to take a large data set and to look for patterns that might inform a correlational structure. And some statisticians would argue a causal structure is possible now. Do you rule out that kind of multivariate study? Do you regard that as an experiment in disguise?

**Thomas:** No, absolutely not. No, well, I mean, I've got nothing against the word experiment. I think. I suppose, you know what I'm talking about there is the appropriation of experiment by particular body of experimenters who - I call it, the Fisher Campbell Stanley tradition of experimenters - who can say or who are claiming to say what works, I've got nothing against, I mean, I don't see the I don't see any valid distinction between quantitative and qualitative work, for example, in in education research, as was really very useful. You know, I think we should be as eclectic as possible in what we do. I expect other criticisms can be drawn of other kinds of research in the way that I've critiqued this particular kind of research. But you know, in all respects, I think I'd encourage eclecticism using using whatever method seems best for the question that you've got. I suppose the back of of what I'm saying really is that we can't say what works from doing this kind of ...we can say what works for me and this situation with these children or young people. Me with my personality, my predilections,

my population of students. But we can't say what works for a whole population based on based on what work studies. I mean, other people would would critique the kinds of research that I'm I particularly promoting. In the article, one of my colleagues talks about qualy-wally's saying, suggesting that qualitative research is essentially pointless and has no value. But no, I would I would think that the kind of thing you talked about a chance is perfectly valid, useful and meaningful.

**Crook:** Yeah, yeah. I think as you're implying, the need for a community conversation about method is really important. And I hope in some small way these conversations, you know, plug into that. There are just two more things I'd like, I'm conscious of time... You've been very generous with it for us, but just two more things I just like to put to you because I'd be really interested to get a response. But one is really relating to what you were just saying about the scale I think of where the researcher is looking, because it's reasonable for a practitioner to regard themselves as a researcher within their own local space as it were. And I'm wondering if there isn't a danger in demonizing the notion of experiment when it is a natural process that you might exercise to clarify something in your own classroom? I mean, you might, for example, you might have a disruptive child and you might think, I wonder if it's the table that they're sitting on with those children and put them somewhere else. And lo and behold, they behave really well. That's a very naive example, but you can see how a teacher might experiment with environmental structures. The ecology of the classroom. And find things really useful. Now we wouldn't want - I put to you - we wouldn't want to discourage an experimental attitude when exercised at a local level.

**Thomas:** Yeah, absolutely. I couldn't agree more. Charles Yeah, I mean, and I've used that kind of experiment myself and in consultation and collaboration with teachers. Any course, one experiment where you've got a baseline and you introduce something, you go back to baseline, you introduce it again and you look to see whether there's any difference between the baseline and the and the implementation of some intervention. I think that's that's absolutely valid and. One can, I mean, you don't necessarily have to say that that that can't be drawn upon by by others, I mean, if people say it might be useful to publish that kind of that kind of finding if they've they've done that sort of experiment with a small E, then others may draw from that and try it out for themselves. What seems to me to be problematic is when you've got this sort of helicopter research, as I've called it somewhere else where, you know, researchers are doing this research

for teachers and dropping it in as some kind of aid, which is, you know, not necessarily meaningful or useful for them. Yeah, certainly doing a controlled experiment or of one kind or another in a particular situation seems to me to be perfectly valid. Yeah.

**Crook:** Okay. Last last prompt from me, Gary. And maybe it's the natural place to conclude, but we're likely to be talking about and talking to the community of research or practitioners and and/or their their own educators. I mean, do you have a thought about the implications of what's said in the paper for how we? Think about undergraduate and postgraduate training in education.

**Thomas:** Well. I suppose the I wasn't principally thinking about that when I wrote the article, except insofar as it's the general, it's the substrate you like it if you like of teacher education. Research is an essential part of teacher education, I think, or learning how to do research well. I think confidence needs to be given back to teachers to do their own kinds of research. And it seems to me that's almost been taken away from them by this assumption that somebody up there is going to be able to tell them what to do. You know, doing your own kinds of practitioner inquiry, it seems to me to be essential that the teachers have confidence in doing that and doing it effectively. So learning about various different research methods, learning about how to to value your experience in developing new kinds of new kinds of knowledge and new kinds of practice seems to me to be essential. And yeah, I mean, one of the problems with this, with the experiment with a capital E is that it's almost the antithesis of that. It's it's saying that there are ways of doing it better. I mean, I suppose this is separate from whether it's actually finding those things out. I don't think it's finding those things out, but the assumption that it might be might disempower teachers from developing their own forms of forms of understanding in their own practice in better ways.

**Crook:** Yeah. I mean, as you've been talking, I think I feel that what's been forming in my own mind is in relation to this question of teacher education is two spaces really one, which is about how the individual practitioner may adopt to kind of critical and analytic attitude towards their own ecology. And then I think we're saying that the discipline of a variety of methods is worth understanding and relating to. But also there is a need, perhaps for practitioners to be critical about the 'what works' messages that might be passed down to education and that your arguments about particularly about aspects are an important break in that kind of wall of skepticism or at least that wall of

critical consideration. Is that fair that we're trying to really cultivate both both senses of a critical awareness?

**Thomas:** Well, I hope so. Yes. That, yeah, people should be critical about about that methodology and be more confident about their own ability to do research and to understand their own situations and their own strengths and and how those combine. Yeah, I mean, there is something a bit dogmatic about, well, not a bit dogmatic, very dogmatic about. About the RCT and the experimental method and the whole talk of gold standards, which I think is almost hypnotized governments and policy makers and funders. Since about 2000, when when medics started talking about evidence based practice and educators thought it would be a good idea to to start talking about evidence based practice as well, not really looking at the history of what that kind of evidence was in education. Yeah, I mean, there is this idea that it is the best and you better do it or you're not doing the best that you possibly can. My next article, it's going to be called gold standards and silver bullets. There aren't any gold standards and there aren't any silver bullets in education as far as I can see. And and I kind of remember the guy who's who's putting together all the various pieces of research that's, you know, the best seller, the Australian guy whose name eludes me for the moment. But that kind of assumption that helicopter research that we can tell you what to do next. I think to be, I don't know, devalued. But you know, people need to understand that it's it doesn't provide everything that it claims it's going to provide.

**Crook:** Ok, I'm going to give you back your time now, go. We'll look out for that forthcoming silver bullet or magic bullet. I really like the metaphor for helicopter research. But there are many things that you've told us that I'm sure we'll give people things think about, and I hope it'll also lead them if they haven't already done so. To read your article so once again, many thanks.