Problems with Generalization in Education Research, Their Consequences, and Their Implications

Jeffrey Pocock
Institute of Education, University of London

Education research is a richly funded enterprise. Millions every year are spent on myriad studies intent on illuminating one of the many areas (assessment, school type, teacher quality, and technology, for instance) that are typically of interest to education researchers. To education research, generalization is fundamental. After all, if research findings cannot be extended to another context or person then institutionalized forms of research lose much if not all of their overriding *raison d'être*. However, in the essay that follows I problematize the very idea of generalization in education research, and discuss the current and future impact, for research, education, and society, of persisting with institutionalized education research where generalization is problematic.

I begin by giving a brief overview of the education research industry before offering a typology of five key forms of generalization to be found in such research. I then present three main problems related to essentialism, methods of reasoning, and evidence that could be said to undermine these five forms of education research generalization, and argue that it is interest and need that must account for the fact that these problems with generalization have not proved insurmountably fatal to the education research enterprise. Lastly, given such problems I discuss whether education research ought to be in the business of generalization at all, set out the potential consequences of ignoring the issues that are raised by problems with generalization, and conclude by describing two futures for education research: one where problems with generalization are addressed, the other where they are not.

On research training courses, education research is typically presented as a variety of methodologies (such as case study or ethnography), methods of data collection (interview or observation, for instance) and analysis (statistical, discourse, and so forth) employed by organizations (providers) in the name of education. In the United Kingdom the largest commissioner of education research is the Department for Education (DfE), but increasingly, either working alone or in collaboration, research is commissioned and/or provided by social enterprises (charities), universities, think tanks (policy institutes), and management consultancies.

The global market for education research is a large one and the amounts spent on it compare favorably with other areas of education and public spending. In the United Kingdom, for example, the amount of government-funded education research has been estimated at £50–60 million (approximately $80–95 million),[1] roughly a tenth of the amount spent in the United Kingdom on post-secondary, non-tertiary education, and two-thirds of the amount spent on civil defense.[2]

A key feature of institutionalized education research is generalization. Etymologically, generalization has its roots in *genus* (or family). As we know, families can

be formed in various ways and this variety is reflected in the plurality of types of generalization that could be said to be found in education research, five of which are outlined below. A typical form of generalization (what will be called generalization$_1$) is that which leads from particulars to general concepts (such as "school" or "pupil"). A second type of generalization (generalization$_2$) is the kind most famously associated with David Hume, and which describes the process of inferring what will happen in future from what has happened in the past.[3] Another form of generalization (generalization$_3$) is statistical or probabilistic generalization that we commonly associate with modern large-scale education research studies.[4] A fourth type of generalization crucial to education research (generalization$_4$) is where we draw conclusions from a specific thing to a more general one.[5] A fifth type of generalization (generalization$_5$) is that of sample-to-population generalization and is another most commonly associated with statistical research.[6]

This typology of different forms of generalization ought to be reasonably familiar to philosophers. For instance, Hume's writings are standard texts on most undergraduate philosophy courses. However, it is unlikely that education researchers, many with backgrounds in statistics and the social sciences, have sufficient awareness of forms of generalization such as the ones typified above and it is unlikely further that education researchers currently have a proper understanding of the inherent difficulties with such forms.

Typically, the problems with education research have revolved around validity and reliability (or variants thereof). If generalization is identified as a problem it is usually a problem of statistical generalization (generalization$_3$, also known as external validity or generalizability). Fundamentally, however, the problem of generalization$_3$ is not seen as fatal to the research enterprise, and commonly logical, methodological, or statistical fixes (such as "average man" or "effect sizes") or alternative strategies (similarity judgements[7], transferability, and fittingness[8], for instance) have been devised with which to address it and above all so that research can continue.

More importantly, in education research other problems with generalization to be described below are rarely if ever explicitly discussed. This lack of explicit discussion can perhaps be explained by the fact that education research training courses and material,[9] which the present author has experience with, do not by convention devote significant time or space to philosophical matters outside of discussion of so-called research paradigms and their ontological and epistemological characteristics, and discussion of what of the paradigmatic smorgasbord a new researcher ought to choose.

It is one belief underlying this essay that if philosophy was awarded greater attention in research training programs, then it is likely that would-be education researchers would come to know more about problems with the types of generalization presented above. For example, researchers might come to know more about the controversies surrounding essentialism and how seriously this problematizes the various types of education research generalization described in this essay.

doi: 10.47925/2013.464

Essentialism is a long-standing argument that the world and its contents that we describe in our language are pre-ordered and distinct, and that what is pre-ordered and distinct are essences (or kinds). Language therefore is a mirror image or representation of the world out there and the distinct essences that comprise it ("human-ness" or "dog-ness," for instance). For education research this means that when using general concepts (generalization$_1$) such as "student" or "teacher" what is being represented or mirrored in our language is the same across research studies. That is to say, when one education research study refers to schools, another completely different study even though researching completely different schools is at the same time still referring to the essence "school" or "school-ness."

Historically, essentialism has been the target of much criticism. Largely it has been rejected on the grounds that essentialism demands too much of language. Rather than language reflecting distinct essences or kinds, it is argued that language instead reflects resemblance between things (such as, properties, qualities, consequences, and processes). The most famous proponent of the resemblance approach is Ludwig Wittgenstein and his family-resemblance concept. For instance, Wittgenstein writes of our tendency or inclination, "to think that there must be something in common to all games, say, and that this common property is the justification for applying the general term 'game' to the various games; whereas games form a *family* the members of which have family likenesses. Some of them have the same nose, others the same eyebrows and others again the same way of walking; and these likenesses overlap."[10]

The idea that instances are subsumed under a general conceptual term because they share a common property (among other properties that are not shared) rather than all instances of a term representing an essence unique to such instances has proven popular in the arts and social sciences. However, difficulties with the resemblance approach to general terms render it unstable. The main objection is that resemblance in property or quality, for instance. is a metaphysical step too far. That is to say, for things to resemble in any way is to assume that what resembles, resembles pre-determinedly. However, this is putting the proverbial cart (*how* something resembles) before the horse (*what* resembles). For example, to put a number of people under a general term such as "student" because they are fifteen years old and in education is to assume that there is a property "fifteen-years-old" and a property "in education," but this rather begs the question as to what if anything these properties ("fifteen-years-old," "in education") actually are.

A second, closely related issue with the resemblance said to be underlying our general concepts is the fact that resemblance seems to contradict an important lesson drawn from evolutionary theory that there are no absolute species or kinds. If this is so, neither, then, are there absolute properties, qualities, and so on and therefore it follows that no two things could resemble each other absolutely in these ways. Indeed, what is perhaps forgotten is that to argue for absolute kinds (or properties, qualities, and so forth) is to argue for something akin to anti-evolutionary Creationism (that is, the idea that in a week the world and its objects were created in a full and distinct form by God).

doi: 10.47925/2013.464

Rather than kinds, properties, qualities, and the products of generalization, one potential, alternative source of resemblance is that of the process of generalization. Many commentators have described generalizations as laws[11] or law-like[12] and perhaps by comparing generalizations to laws we can make resemblance work.

Given the discussion so far, however, it would appear that any attempt to defend the resemblance between laws and generalizations cannot even get off the ground. This is due to processes just as much as properties or qualities being subject to the same metaphysical objections (thus, the cart-horse problem) of presupposing that which we hope to demonstrate.

Moreover, even if we were to put aside the fact that claiming two instances resemble each other because of the processes that led to them relies on a dubious, anti-Darwinian essentialist metaphysic, it still may not be desirable to do so. This is if we take account of the well-known problems (to philosophers, that is) with various processes of generalization (or general reasoning).

Typically, there are said to be three distinct processes of generalization known as induction, deduction, and abduction. Inductive discovery and justification is exemplified by the five types of generalization above. General concepts (generalization$_1$) can be said to be derived (discovered) from a number of particular instances, but so can they be justified inductively. The other four generalizations (past-to-future, statistical, specific-to-general, sample-to-population) are also all known as inductive generalizations. Deduction meanwhile is known classically as a process of inferring consequences from a set of premises. It is also known as hypothetico-deductive method and is a method that features prominently in statistical research. Abduction (or inference to the best explanation) lastly is known as a process of identifying the most likely explanation (set of premises) for an event (consequence).

Over centuries of discussion in the literature these three processes of reasoning either individually or in some kind of combination have won the support of various personages. Of particular note are the writings of Aristotle on deduction,[13] Francis Bacon on induction,[14] and Charles Sanders Peirce on abduction.[15] However, what these and other various personages have not been able to do is convincingly address the inherent fallibilities of their favored process (or processes) of reasoning. For instance, as with induction and deduction, abduction is criticized for being circular (the so-called bootstrapping problem) — that abduction is an infallible/best method of generalization is a generalization that has been arrived at abductively — or it is self-defeating — that abduction is an infallible/best method of generalization is a generalization that has been established inductively or deductively. The most telling criticism of deduction is that false premises can have true consequences (all men are mortal; Cleopatra is a man; therefore Cleopatra is mortal). Induction meanwhile was regarded as a myth by Karl Popper because of the alleged implausibility both of a person interacting with the world in possession of a pre-theoretical blank-slate and of inductive generalizations somehow emerging from such an original position.[16] In addition, as well as being the inspiration for Popper and highlighting the bootstrapping problem, Hume identified a missing link between inductive premise and consequence.[17]

doi: 10.47925/2013.464

How is it, he asked, that we assume that the future will be like the past if there is nothing inherent in the past event to tell us so? What Hume concluded then was that our general inductive claims are only as good as the last observation.

A third major problem with generalization is the problem of underdetermination[18] known well to philosophers and sociologists of science. Underdetermination challenges the idea that it is the world out there in its distinctive, essential form that is the final and definitive arbiter of a theoretical test. Why this is so is because it is said firstly that no theory can be tested in isolation and said secondly that there can be alternative, even vastly different theories that compare as well as the one that is favored and tested. It is then not known for sure what part of our theoretical web is being tested, nor is it known whether the prevailing general hypothesis or theory, which we think we are testing, is the right one. In effect, the judgement as to whether a general theory has been fatally wounded or not is down to us rather than determined by the evidence.

In education research, we can see how underdetermined general theories are, if, for example, we consider the very many suggestions made as to what contributes most to academic performance (class size, teacher quality, or family income, for instance). In not one case can it be said that the evidence wholly determines one of these general hypotheses.

Given the difficulties with essentialism, processes of reasoning and evidence, how is it that generalization remains so central to education research? One reason why generalization in education research continues seemingly oblivious to these problems with generalization is the lack of philosophical content in research training courses described above. But it can perhaps also be attributed to what might be termed as interest and need.

Attributing generalization in education research to interest and need may strike some readers as reminiscent of the Sociology of Scientific Knowledge (SSK) program. Taking their cue from Robert K. Merton and Thomas Kuhn, in particular, SSK contributors (including Barry Barnes, David Bloor, and Harry Collins) conclude that logical positivists, Popper, and others (the so-called First Wave of Science Studies) have failed in their ambition to uncover a perfect, fool-proof method of generalization and that career ambitions, workplace politics, money, and so on, account for the way that science actually works.

Although broadly sharing this view, there are distinct differences between SSK and the approach taken in this essay. Firstly, I acknowledge and so seek to avoid a tendency in SSK to replace imperfect rationality with an alternative but just-as-metaphysical set of interests and needs.

I wish to make clear that I do not stake a claim for a world necessarily *driven* by interests and needs but rather suggest that we understand what education researchers do in this way, as that which *prima facie* research would appear dependent on for success (process, essence, evidence) is seemingly not up to the philosophical job. It is then not that interests and need are explanatory but more that process, essence, and evidence are not.

doi: 10.47925/2013.464

One good reason for not staking a metaphysical claim in this way is that to do so subjects interest and need to the same problems of underdetermination and so on that bedevil that which it hopes to replace. In SSK, responses to this quandary have been, for example, to recognize randomness and chance factors as determining the generalizations made by researchers,[19] but even this still presupposes a particular metaphysic (that is, it presupposes a world of chance).

My argument then is a negative one: processes of generalization, essentialism, and evidence are inherently fallible drivers of generalization in science (and education research), so as a consequence it must be something else that is responsible for the persistence of research generalization, and what better way to describe it than interest and need. In other words, interest and need are what methods of reasoning, evidence, and essentialism are not.

Another way that this essay differs from SSK is that I am acutely aware of the inconsistency of self-excepting myself from the arguments that are being made. In arguing that methods of flawless reasoning, accurate linguistic representation of real essences or kinds, and conclusive evidence are perhaps unattainable, it is important to recognize that this applies to this essay too. For instance, in the case of underdetermination of evidence, other general theories (concerning intuition, for example) can reasonably be offered to account for the ways in which education researchers generalize.

That this essay is also subject to the problems with generalization raises the prospect that *no one* is able to overcome such problems. Fundamentally, if we are all generalizers in the same, imperfect way, this means that neither researchers nor anyone else can rely on flawless generalization to legitimize their decision-making authority. For some in SSK[20] this has meant that the problem of the continued legitimacy of what scientists do has been replaced by the problem of extension or how to extend the moniker of "scientific expert" to non-expert others. They ask: if we are equal in our generalizing abilities how then can we effectively distinguish between expert and non-expert, scientist and quack?

This question of how to distinguish between expert and non-expert (the problem of extension) will not be explored any further, however. This is because of the last and final difference between the SSK agenda and mine. Unlike the relative silence in SSK, I seek to elucidate an issue more pressing than addressing the problem of extension: that of highlighting the political and social consequences that are a result of the faux-objectivity that the rhetoric of infallible method, accurate essentialist representation, and conclusive evidence contribute to. The overriding matter then is not that the problem of legitimacy has been replaced by that of extension, but that the problem of legitimacy still persists. Collins and others in the SSK movement perhaps err in assuming that as the philosophical case is so conclusive, we can somehow move on from the problem of legitimacy to another and newer problem (of extension). This neglects the fact that problems with generalization and so of legitimacy are still rarely discussed in education research circles. What is needed then is to spell out the potentially damaging consequences of persisting in this unphilosophical way.

doi: 10.47925/2013.464

If we accept that there are problems with generalization but that we are all subject to them when we generalize and that currently education research appears to act *in spite of* such problems, then this has a number of consequences. One such practical problem pertains to the amount of time it takes to commission and secure funding for research, to conduct it, analyse the data, report findings, and see these findings implemented in policy. For example, consider the numerous years that each Programme for International Student Assessment (PISA) cycle takes from assessment to implementation of policy in light of the results.

Of course, it may be argued that even with lengthy lead times, researchers with their special methods, grasp of essences, and/or access to the best evidence, are better placed to draw general conclusions about schools, teachers, and so on. However, as has been discussed, the ability to generalize may well be shared equally by all but also equally fallible for all. If the ability to generalize is a flawed one, and one that would not appear to be the preserve of education researchers but of all able-minded people, then whether professional education researchers or, for example, classroom teachers generalize makes no difference.

Thinking in terms of efficient spending of public funds, if it makes no significant difference who generalizes, it is better that the least expensive option is chosen, and this option would seem to be allowing teachers on the ground to research rather than meeting the considerable expense of other professional researchers doing it. The £50–60 million, for example, spent on UK education research in other words could be better spent.

A third major consequence of ignoring the problems with generalization in education research pertains to testing and once formulated the very real possibility that generalizations are not tested, or not modified in line with further testing. John Dewey referred to generalizations that were not tested as inert and dead abstractions.[21] It could be argued that this reluctance to test generalizations, which to Dewey renders them inert and dead, is a result of essentialist assumptions about the world that, as noted above, are not normally broached in discussion of problems with generalization in research. If the world is characterized by unchanging but distinct essences, a single generalization in research suffices and suffices as the last word. After all, why keep generalizing and testing those generalizations in classrooms and schools, for example, if the world is composed of fixed and timeless kinds? Repeat testing and modification of generalizations would seem to an essentialist a waste of time and money.

A corollary of this reluctance to conduct further tests — perhaps because of adherence to essentialism, but also perhaps because of the assumption that the evidence is definitive or method of reasoning infallible — is the encouragement given to the entrenchment of social and educational institutions that exists to satisfy the demands of specific generalizations.[22] Such institutions would seem to have an interest in general claims that they depend on surviving. The survival of these pivotal generalizations, the truth of which such institutions rely on for future funds and support, suggests testing (thus, criticism) of such general claims is and will not be appreciated, nor in light of ostensibly contrary evidence will any demands to limit

doi: 10.47925/2013.464

these claims. Such dogmatic intolerance to criticism could lead to its eventual suppression, by, or encouraged by, such social and educational institutions, especially where such institutions are close to the corridors of power. There is then a very real threat of authoritarianism where an un-philosophical and dogmatic approach to generalization is adopted.

If we need to treat generalization with more care given these problems with it, what then for the future? For one thing, education research training programmes ought to devote a good deal more time to philosophy, and philosophy of science in particular. Although in my own case as teacher and trainee researcher, I had my suspicions that generalization was a problem worth investigating, it was not until in my own time I had got to grips with the philosophical literature that I could make explicit what my suspicions were made of. Philosophy offers a rich vocabulary with which to articulate these nagging doubts about institutionalized education research. By encouraging researchers to engage with the philosophical literature on generalization, it will enable them to make a better decision as to whether they ought to conduct the type of research that they intend to. They may even decide that they ought not to conduct institutionalized research at all and ought to become teachers instead.

Another implication of this discussion on the problems with generalization is that such problems are not going to go away anytime soon. How then should education research commissioners proceed? There would seem to be two very different futures. If our interest in doing education research is to ensure that the general conclusions drawn from it are as current as possible, then deinstitutionalized education research led by those on the ground (teachers) would seem more appropriate. After all, the literature suggests that those on the ground are no more or less skilled at generalizing then those who are not, and teachers rather than professional researchers would seem to have value-for-money and lead times on their side.

However, if there is no interest in keeping generalizations current because discussion regarding these problems with generalization remain off researcher training courses or because of the threat such updating of general conclusions will pose to well-established social and educational institutions and their profitable public campaigns, then large-scale, time-consuming, and professional-led research projects will continue to be funded. The status-quo with its authoritarian tendencies will remain and the problems with generalization will continue to be the elephant in the (class)room.

---

1. David H. Hargreaves, *Teaching as a Research-Based Profession: Possibilities and Prospects, Teacher Training Agency 1996 Annual Lecture* (London: TTA, 1996), 1.

2. HM Treasury and National Statistics, *Public Expenditure Statistical Analyses (PESA) 2012* (London: TSO, 2012).

3. This form of generalization underpins pilot studies such as NatCen Social Research, Institute for Fiscal Studies, and Bryson Purdon Social Research, *Implementing the Free School Meals Pilot* (London: DfE, 2012).

doi: 10.47925/2013.464

4. "Meta-analyses suggest an average effect size of about 0.4, indicating that pupils might make about 4 or 5 months progress during an intensive programme (of one-to-one tuition)." Steve Higgins, Dimitra Kokotsaki, and Robert Coe, *The Sutton Trust-EEF Teaching and Learning Toolkit* (London: Sutton Trust-EEF, 2012), 18.

5. OECD Programme for International Student Assessment (PISA) tests of reading, maths and science are meant to represent a student's ability "to meet real-life challenges" or the "challenges of today's knowledge societies." OECD, *PISA 2009 Technical Report* (Paris: OECD Publishing, 2012), 22.

6. "Around 470,000 students completed the (PISA) assessment in 2009, representing about 26 million 15-year-olds in the schools of the 65 participating countries and economies." OECD, *PISA 2009 Results: What Students Know and Can Do Student Performance in Reading, Mathematics and Science (Volume I): Student Performance in Reading, Mathematics and Science*, vol. 1 (Paris: OECD Publishing, 2011), 20.

7. Colin W. Evers and Echo H. Wu, "On Generalising from Single Case Studies: Epistemological Reflections," *Journal of Philosophy of Education* 40, no. 4 (2006): 511–526.

8. Egon G. Guba and Yvonna S. Lincoln, *Naturalistic Inquiry* (Beverly Hills, CA: SAGE Publications, 1985).

9. See, for example, Norman Blaikie, *Approaches to Social Enquiry: Advancing Knowledge*, 2nd ed. (Cambridge, UK: Polity, 2007).

10. Ludwig Wittgenstein, *Preliminary Studies for the Philosophical Investigations Generally Known as the Blue and Brown Books*, 2nd ed. (Oxford: Blackwell, 1969), 17.

11. See, for example, Hans Reichenbach, *The Rise of Scientific Philosophy* (Berkeley, CA: University of California Press, 1951).

12. See, for example, Alasdair MacIntyre, *After Virtue: A Study in Moral Theory* (London: Duckworth, 1981).

13. Aristotle, *Prior Analytics*, ed. Robin Smith (Indianapolis, IN: Hackett, 1989).

14. Francis Bacon, *Advancement of Learning and Novum Organum*, ed. James Edwin Creighton (New York: The Colonial Press, 1900).

15. Charles Sanders Peirce, *Collected Papers of Charles Sanders Peirce*, eds. Charles Hartshorne and Paul Weiss (Cambridge, MA: Belknap Press of Harvard University Press, 1974).

16. Karl R. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (London: Taylor & Francis, 2002).

17. David Hume, *A Treatise of Human Nature*, eds. David Fate Norton and Mary J. Norton (Oxford, UK: Oxford University Press, 2000).

18. Pierre Duhem, *The Aim and Structure of Physical Theory*, trans. Philip Wiener (Princeton, NJ: Princeton University Press, 1991); and W.V.O. Quine, "Main Trends in Recent Philosophy: Two Dogmas of Empiricism," *The Philosophical Review* 60, no. 1 (1951): 20–43.

19. See, for example, David Bloor, *Knowledge and Social Imagery* (Chicago: University of Chicago Press, 1991).

20. See, for example, Harry Collins and Robert Evans, "The Third Wave of Science Studies: Studies of Expertise and Experience," *Social Studies of Science* 32, no. 2 (2002): 235–296.

21. John Dewey, *How We Think* (Boston: D.C. Heath, 1933).

22. John Dewey, *Freedom and Culture* (New York: G.P. Putnam's Sons, 1939).

doi: 10.47925/2013.464