Paul Smeyers
Marc Depaepe
Editors

EDUCATIONAL RESEARCH 5

Educational Research:

The Ethics and Aesthetics of Statistics

Springer

Educational Research: The Ethics and Aesthetics of Statistics

Educational Research

VOLUME 5

Aims & Scope

Freedom of inquiry in educational research can no longer be taken for granted. Narrow definitions of what constitutes 'scientific' research, funding criteria that enforce particular research methods, and policy decision processes that ignore any research that is not narrowly utilitarian, in many countries, create a context that discourages scholarship of a more speculative, exploratory, or critical sort.

In this series, internationally leading scholars in *philosophy and history of education* engage in discourse that is sophisticated and nuanced for understanding contemporary debates. Thus social research, and therefore educational research, is again focused on the distinctive nature of what it studies: a social activity where questions of meaning and value must be addressed, and where interpretation and judgment play a crucial role.

This educational research takes into account the historical and cultural context and brings clarity to what actually constitutes science in this area. The timely issues that are addressed in this series bear witness to the belief that educational theory cannot help but go beyond a limited conception of empirical educational research to provide a real understanding of education as a human practice. They surpass the rather simple cause-and effect rhetoric and thus transgress the picture of performativity that currently keeps much of the talk about education captive. The authors are united in the belief that 'there is a place within the social sciences in general', and within the discipline of education in particular, for 'foundational' approaches that enable the systematic study of educational practice from a discipline-orientated approach.

Paul Smeyers · Marc Depaepe
Editors

Educational Research: The Ethics and Aesthetics of Statistics

Springer

Editors
Prof. Paul Smeyers
Universiteit Gent
Faculteit Psychologie en Pedagogische Wetenschappen
Dunantlaan 1
9000 Gent
Belgium
paul.smeyers@ped.kuleuven.be

Prof. Marc Depaepe
Katholieke Universiteit Leuven
Campus Kortrijk
Subfaculteit Psychologie en Pedagogische Wetenschappen
Sabbelaan 53
8500 Kortrijk
Belgium
marc.depaepe@kuleuven-kortrijk.be

ISBN 978-90-481-9872-6 e-ISBN 978-90-481-9873-3
DOI 10.1007/978-90-481-9873-3
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010936458

Earlier Volumes in this Series

Educational Research: Why 'What Works' Doesn't Work
Paul Smeyers and Marc Depaepe
2006, VI, 195 p., Hardcover
ISBN 978-1-4020-5307-8

Educational Research: Networks and Technologies
Paul Smeyers and Marc Depaepe
2007, VI, 228 p., Hardcover
ISBN 978-1-4020-6612-2

Educational Research: the Educationalization of Social Problems
Paul Smeyers and Marc Depaepe
2008, VI, 247 p., Hardcover
ISBN 978-1-4020-9722-5

Educational Research: Proofs, Arguments, and Other Reasonings
Paul Smeyers and Marc Depaepe
2009, VI, 199 p., Hardcove
ISBN 978-90-481-3248-5

Contents

Chapter 1
Representation or Hard Evidence? The Use of Statistics in Education and Educational Research

Paul Smeyers and Marc Depaepe

> *The quantitative and antianthropocentric orientation of natural sciences from Galileo on forced an unpleasant dilemma on the humane sciences: either assume a lax scientific system in order to attain noteworthy results, or assume a meticulous, scientific one to achieve results of scant significance.*
>
> (Ginzburg, 1989, p. 124)

A canonical text of the history of science, more in particular of educational research, reads as follows: 'One cannot understand the history of education in the United States during the twentieth century unless one realizes that Edward L. Thorndike won and John Dewey lost' (Lagemann, 2000, p. xi). Apart from whether or not one agrees with this bold claim (see, among others, Depaepe, 2010; Gibboney, 2006; Tomlinson, 1997), one has to admit that the kind of research that uses quantitative, i.e. statistical techniques, has gained most prestige in the 20th century (see, among others, Depaepe, 1993; Wooldridge, 1994; Richardson & Johanningmeier, 1997; Porter & Ross, 2003; Johanningmeier & Richardson, 2008). Various often interrelated factors are responsible for this, such as the belief in and the acceptance of the assumptions of positivism, the institutional growth of the educational market, the so-called scientisation of educational research, the professionalisation and academisation of the training of education(al)ists, the supremacy of meritocratic values in modern societies and the constant need to legitimate these by 'objective' and 'neutral' research. Unlike his colleague Dewey, with whom he worked for more than 40 years at the renowned Teachers College, University of Columbia, New York, Thorndike embraced this 'trendy direction' of educational research. In 1968 Thorndike's biographer admiringly described him as *the sane positivist* (Jonçich, 1968). As a 'cult figure' Thorndike was the sign of the 'new' world with which the old continent could not keep pace: '... while Europeans were exploring the subjective and personal dimensions of experiences – using the eyes and insights of Bergson, Freud and Van Gogh – Americans are keeping

P. Smeyers (✉)
Ghent University, Ghent, Belgium

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_1,

their art representational, their novels realistic, making their philosophy empirical, their historiography scientific, and above all, their psychology behavioral' (Jonçich, 1968, p. 55).

The use of methods of testing and statistics was at the core of this success story, a story based on the unshakable belief that everything can be measured. William McCall (1922) – the residing statistician at Teachers College – immortalised this unbridled trust in quantification with the well-known assumptions (1) 'whatever exists at all, exists in some amount . . .'; (2) 'anything that exists in amount can be measured . . .'; (3) 'measurement in education is in general the same as measurement in the physical sciences (. . .)' (William McCall, 1922, pp. 3–5). On the one hand these assumptions relied to a large extent on Thorndike's educational psychology; on the other hand, once they were made explicit, Thorndike and his followers were eager to adopt them in order to further justify the way they saw research; as the antidote against all societal evils (see, e.g. Travers, 1983).

This is not the first time that the *Research Community* '*Philosophy and history of the discipline of education*',[1] established by the Research Foundation Flanders FWO, Belgium (Fonds voor Wetenschappelijk Onderzoek – Vlaanderen), addresses an area that is paradigmatic for educational research. In both the first (1999–2003) and second (2000–2008) periods, which focused on *Evaluation and evolution of the criteria for educational research*, various positions were scrutinised (see Smeyers & Depaepe, 2003, 2006, and Smeyers, 2008). In the present (third) 5-year period of this *Research Community* (2009–2013), the overall interest is *Faces and spaces of educational research*, which is divided into four subthemes (respectively addressed during the conference in 2009, 2010, 2011 and 2012): the ethics and aesthetics of statistics; the attraction of psychology; institutional space; designs, material culture and the representation of educational research.

The chapters published in this volume were first presented at the 2009 *Research Community* conference in Leuven. Scholars from philosophy and history of education (some of whom are particularly interested in history and philosophy of science) combine their efforts to study statistics as part of both the academic discipline of education and the broader educational context. Statistics are (still) everywhere. Their power and undoubted efficacy in many areas has given rise to the same faith in measurement and metrics. The more statistics we gather, the more we will know. Their use carries with it a number of presuppositions: that reality is being represented, that it can be controlled and the risks can therefore be managed. As case studies, the chapters interpret the ethics and aesthetics of statistics in terms of representation, visualisation and accessibility, the appeal of 'simplicity', of technical languages, numbers, diagrams and pictures, and pay attention to their connection with action plans. At first sight, some of the observations and arguments made by the contributors may give the reader the impression that statistics has only negative connotations and that it should be banned from educational research altogether as its contributions are dubious to say the least, and moreover as one tends to present these results as hard evidence.

This is not what the contributors are trying to say and this is a distorted picture embracing a polarisation that should be opposed. Statistics should neither be seen as

the golden or only road we can follow to understand educational reality nor should their importance be disregarded when the issues that are studied require such an approach. For instance, if one is interested in a phenomenon such as 'bullying' in primary schools, one evidently wants to know how many cases of bullying have been registered. This can be specified further for particular subgroups such as boys and girls, according to age, ethnicity, various living conditions, and so on and so forth. Now it goes without saying that to have an informed estimate of the frequency of the occurrence of a particular problem (as detailed as this can be) is quite essential in educational contexts. Policy needs to take this into account, as it can be an element in the process of determining how relevant the problem is. This not only has implications for what should be done (by whom and at what level), but also for the kind of research that should be carried out (not to mention the quantity of researchers and research funding that should be mobilised). And it is evident that questions about, for example, the relationship between different phenomena in large populations (To what extent is underachievement in school linked to gender, ethnicity, or social class? Which teaching approaches are related to high achievement in reading tests?) – i.e. correlational studies – would certainly require a sample larger than $n = 1$ (see Chapter 6 by Bridges, this volume). Such studies could also point more precisely to what ought to be addressed in this research. They could alert scholars to phenomena that may have gone unnoticed (an example from a different area is, for instance, the prevalence of certain types of cancer in particular geographical areas). Moreover, a 'thick description' of a school or community of the kind normally associated with qualitative research might and perhaps should include quantitative information about the social class make up of the school, the distribution of test scores of pupils, staff student rations, etc. There is, we think, a legitimate place for different kinds of quantitative research within educational research or, even more broadly, within the academic discipline of education (see Chapter 11 by Smeyers, this volume). What we want to underscore here, however, is the more general criticism that can rightly be raised against the exclusive use of randomised field trials and, more generally, experimental or quasi-experimental approaches (by some labelled as the 'Gold Standard'), approaches that have often been used in educational policy to justify certain interventions. There is a plethora of criticism internal to the use of particular statistical techniques, but there is generally a lack of external criticism that takes into account the overall picture of what education and child-rearing should be about. As many of the chapters in this volume show, a crucial element is the way problems are conceptualised. This has far-reaching consequences for the kind of decisions that are taken on the basis of research. This straightforward point is often forgotten when people look at what research 'tells' us.

The contributors to this volume, who are all working in philosophy and/or history of education and who are all particularly interested in philosophical and/or historical aspects of the discipline of education, will point to the lack of appreciation of the relevance of *the concept(s)* and raise questions concerning the application of the research findings (in other words demand attention to the crucial importance of contextualisation). However, even in their own areas, they do not necessarily doubt the contribution statistics can make to particular research questions. This holds even

more for other areas of educational research as long as statistics are not seen as a goal in themselves but as a tool to acquire understanding. Whereas such an approach is gradually winning acceptance within the sociology of knowledge, it is rarely adopted in the historical study of educational institutes including research institutes and universities (see Depaepe, 2010). In our opinion, it is important that this domain should start with statistics so as to acquire a better understanding of the networks that have played a part in the development of various disciplinary matrixes. For example, since the 1970s many educational journals have been the object of qualitative educational research, yet avenues for exploring this further from a quantitative point of view have rarely been mapped. Recently, Tenorth (2010) tried to identify the part played in the 20th century by empirical research in the area of child-rearing and education. Only if such networks of producers of educational knowledge (as well as of those of the gatekeepers and consumers of knowledge) are uncovered, will the cartography or 'social geography' of the discipline of empirical or experimental educational research be shaped from a history of science perspective.

This book explores what made educational researchers dependent on statistics. It deals with topics such as the use of statistics for measuring the prevalence of maltreatment of children, European citizenship and evidence-based happiness, irregular migrants, and for university expansion. The book also explores the drive to boost statistics, which finds its voice in policy initiatives that become slogans and looks at how public opinion polls are used to rationalise political decision making. It questions whether a more limited and modest use can be made of statistics which does not deflect attention away from education's core business and which does not destroy the local practical knowledge that makes the educational area function effectively. The attempts to answer these questions find their expression in 13 case studies from the stance of philosophy and/or history of the discipline of education.

In Chapter 2 David Labaree, explores the historical and sociological elements that have made educational researchers dependent on statistics – as a mechanism to shore up their credibility, enhance their scholarly standing and increase their influence in the realm of educational policy. He begins by tracing the routes of the urge to quantify within the mentality of measurement that arose in medieval Europe and then explores the factors that have pressured disciplines and professions over the years to incorporate the language of mathematics into their discourses. In particular, this pattern has been prominent for domains of knowledge and professional endeavour whose prestige is modest, whose credibility is questionable, whose professional boundaries are weak and whose knowledge orientation is applied. The chapter shows that educational research as a domain – with its focus on a radically soft and thoroughly applied form of knowledge and its low academic standing – fits these criteria to a tee. It then examines two kinds of problem that derive from educational researchers' seduction by the quantitative turn. One is that this approach to educational questions deflects attention away from many of the most important issues in the field, which are not easily reduced to standardised quanta. Another is that by adopting this rationalised, quantified, abstracted, statist and reductionist vision of education, policy-makers risk imposing reforms that will destroy the local practical knowledge that makes the ecology of the individual classroom function

effectively. Quantification, Labaree argues, may be useful for the professional interests of educational researchers but it can be devastating in its consequences for school and society.

In Chapter 3, Marc Depaepe deals with a report published in 1964 where several Flemish intellectuals argued that, 'in order to enhance the quality of university education as much as possible', undergraduate university campuses had to spread out geographically. This would not only reduce the deficiency of certain areas as regards university recruitment but also the social backwardness that was accompanied by this deficiency. Anyone who observes the development of university education over the past 50 years will quickly conclude that this did not occur as directly as the proponents of the 'dissemination of undergraduate education' had hoped. This 'university expansion' – which was accomplished in two phases: in 1965 and in 1971 – was accompanied by a spectacular growth in the number of university students, though that is not to say that this was synonymous with 'democratisation'. Not only did the 'massification' of university education appear to be due more to the success of the traditional campuses, but sociological research also raised doubts about the intended social effect, since children of the less educated made much less use of the university 'expansion' than did the children of the more educated. Historians do not have much difficulty with the plausibility of this last conclusion. That education, as a social institution, primarily bears a 'bourgeois' and 'meritocratic' character has long been recognised in the history of education. Still, the conclusions of the sociological research referred to cannot claim, historically, to be much more than hypothetical and/or heuristic. They are interesting preliminary studies that, with regard to the problem of the democratisation of university education, need both a cultural historical interpretation and more pertinent (*in casu*, primary) source material – a thesis that is further elaborated in the chapter.

In Chapter 4, Jeroen Dekker refers to the fact that after the publication in 1962 of the article on the *Battered Child Syndrome* by the American medical doctors C. H. Kempe, F. N. Silverman and their colleagues (Kempe, 1962), numerous studies were published on abused and neglected children. Moreover, an increasing institutional and legal framework of diagnosing and preventing child maltreatment was set up in many countries in the Western world. In this chapter, the question that is asked concerns whether this increasing world-wide interest in the maltreatment of children resulted in a major diminution of child maltreatment or not. Although the hypothesis that maltreatment of children was diminishing, at least in the economically prosperous Western world, looks strong, the opposite seems to be true when looking more thoroughly into the information available. This chapter focuses on statistical studies on the prevalence of maltreatment of children in the Western world paying special attention to the USA and the Netherlands. Part of the answer to the question relates to the multiplier effect of three phenomena: a broader definition of child maltreatment since the 1970s; the impact of internationally accepted children's rights by prescribing criteria for good parenthood and for child protection; finally, the preference of policy-makers for clear figures or for the aesthetics of statistics, in developing child protection policies – and this notwithstanding the fact that they are confronted with sometimes contradictory figures.

The chapter concludes by maintaining that these three factors vastly contributed to the increase of the reported prevalence of child maltreatment in the Western world.

In Chapter 5, Norbert Grube claims that social sciences create phenomena. Opinion polls construe the formation of public opinion by selective questions within questionnaires. John Dewey saw social psychological analyses as a means to transfer American mass society into the Great Community, whereas Walter Lippmann wanted to put social sciences into the service of the government. The founding fathers of polls in the USA, especially George H. Gallup, referred implicitly to Dewey and regarded polls as a science that would strengthen democracy. This sentiment echoed in the sentiments of Elisabeth Noelle-Neumann and her husband Erich Peter Neumann, the founder of the Allensbach Institute of public opinion research in 1947, which is a central focus of this chapter. Polls do not only aim to establish and improve democracy. They also aim to create national conformity. The nation stands in the focus of polls, especially if national governments are the customers of pollsters. The presentation of data suggests clarity within the findings along with the possibility of reliable predictions, certainty, comparison and competition. The simplification of data in a few charts shall help to rationalise the political decision-making process for a preventative population policy. At the same time, the graphical simplification should support notions about the social and national body that are based on ideals of unity. Furthermore so-called clear diagrams and findings of applied social surveys should be the convincing starting point for political campaigns and for people's instruction. The chapter reveals different models of dualistic questions and dichotomous models of graphical designs. This dualism does not only divide the people into two camps but establishes a new national narrative. The exact and suggestive presentation of findings led to several attempts to instruct and convince the West Germans of the veracity of capitalism, the democratic system and the European unification. But because polls often present the respondents as incomplete human beings, it is questionable if polls could redeem the aim of preventative governmental policy or whether they could engender a democratic community based on communication.

In Chapter 6, David Bridges argues that individual cases can and do have enormous rhetorical and motivational power in public policy debate: at the same time contemporary discussion of 'evidence-based policy' tends to push the individual case study to the margins of policy-makers' interest, which is focussed instead on population studies and randomised controlled trials in which large numbers provide the appearance of validity and a sense of confidence in the results. This chapter seeks to examine the role of the single case (and by extension small numbers of cases) in informing educational policy and practice, asking how, why and under what conditions *should* educators pay attention to such research. It begins by looking at the reasons for the predilection for large numbers in quantitative research, but also the role of the single case in these research traditions in challenging generalisations and inviting closer examination of the particular context in which apparently aberrational results are observed. This chapter then observes the significance of the 'case' in a variety of academic communities – in psychology and the development

of psychoanalytic theory; in ethnography; in auto/biography and life history; in history itself and in law. It identifies the main characteristics of case study, at least in the forms that are most familiar to educational enquiry. A central part of the chapter examines some of the arguments around the possibility of generalising from individual cases starting with the view that educational conditions are so locally defined by time and social context that such generalisation (from any form of research) is doomed from the start. It considers three particular functions of the case study: (i) as a source of conjecture (and grounded theory) and refutation; (ii) as a basis for understanding one particularity by reference to another (without any attempt to generalise); and (iii) as providing a vicarious form of experience and making a contribution to the reader's practical wisdom. A final section of the chapter considers the sense in which the conduct of case study research resembles scientific or artistic approaches to understanding. Following MacDonald and Walker it concludes with the notion that 'case study is the way of the artist, who achieves greatness when, through the portrayal of a single instance locked in time and circumstance, he communicates enduring truths about the human condition'.

In Chapter 7, Elias Hemelsoet argues that for various reasons, irregular migration has become a more frequent phenomenon during the last decades. Humanitarian and social problems related to this fast-growing group of people give a boost to political discussion and subsequently to scientific research. Both politicians and scientists want to 'grab' the situation and acquire an overview of what is happening. Estimating the number of irregular migrants in a country is in most cases the base to deal with emerging problems. A large number of different methods are used to make these estimations. Unfortunately, these bring forth rather feeble results, which is partially a consequence of a large margin of error. Nevertheless, this does not stop people from basing policy with far-reaching implications on these 'findings'. The chapter focuses on a methodological question concerning estimations of irregular migrant numbers, i.e. the problematic character of conceptualisations. This problem will be treated in two ways. First, different definitions of people without legal residence lead to specific ways of conceptualising the problem. This is problematic as different statistical outcomes are generated which will lead to multifarious conclusions and recommendations. Second, there is the intrinsic danger of reducing the complexity of a large amount of data to a limited number of variables by taking away the possibility of making distinctions that are often desirable (which unavoidably implies a tendency towards homogenisation). In the case of irregular migrants, their illegal residence is the benchmark of categorisation. The level of diversity regarding the particular circumstances of these people is often ignored as policy treats them as a monolithic group. To conclude, Hemelsoet considers the extent to which questions put forward by policy-makers and researchers 'make sense' when it comes to dealing with the problems at stake. Why are these questions so attractive and apparently self-evidently valid?

In Chapter 8, Naomi Hodgson focuses on the objects of statistical analysis that provide a current focus for measurement and policy-making in Europe and thereby constitute key indicators according to which states, institutions and individuals become measurable, comparable and governable. She focuses, in particular, on the

concern with happiness and well-being and how this relates to the construction of 'active citizenship'. With reference to policy literature relating to the development of measures of active citizenship, European surveys addressing citizens in terms of their happiness and well-being, and school curriculum recommendations on 'Social and Emotional Aspects of Learning', the chapter seeks to illustrate how we are asked to account for ourselves according to a particular language of active citizenship. It is suggested that attention to the technologies according to which statistical knowledge operates, in light of the increasingly complex relationship between the governmental and the private (commercial) sources of knowledge production, is necessary for understanding how we are made subjects, and what 'citizenship' means, today.

In Chapter 9 Ulrike Stadler-Altmann and Edwin Keiner focus on the contexts empirical educational research knowledge is embedded in. They consider how such knowledge becomes the object of processes of de- and re-contextualisation and functions in accordance with the expectations of various social groups. On the one hand, they identify an aesthetics of educational research knowledge and the rhetorically persuasive power of figures and graphic accounts (in other words an aesthetics of the way 'answers' are provided). On the other hand, they draw attention to an aesthetics of deconstructive scepticism and of epistemological relativism, i.e. an aesthetics of raising questions. The combination of both aesthetic forms, the provocation of both giving answers and raising questions in a research presentation formatted in a particular way, is related to two further aspects of aesthetics that the chapter goes on to address. They describe these aspects as a 'more rhetorical' and a 'more epistemological' aesthetics. Their analysis deals with (a) (oral) presentations of educational research projects which are examined in a micro-analytical way and (b) (written) publications addressing the knowledge society or educational research outcomes based on large-scale assessment studies, especially those carried out by the OECD. These written publications are approached from a macro-analytical perspective. The chapter offers a comparison of a poem with a presentation of educational research. It focuses on the role figures and statistics play in the rhetorical aspect of an exposition of research results. This amounts to a 'grammar' of presentation.

In Chapter 10, Jean Paul Van Bendegem, Karen François and Kathleen Coessens interpret the ethics and aesthetics of statistics in terms of representation, visualisation and accessibility. A specific case study is briefly examined, namely the development of ISOTYPE (International System Of TYpographic Picture Education) by Otto Neurath, as an attempt to represent statistical data in such a way that any citizen could have access to it. They extend this case study into a set of observations concerning the use of diagrams and pictures in, for example, mathematical reasoning and the semiotic perspective that allows one to connect Neurath's work to, for example, C. S. Peirce's approach. In the second part of the chapter they move from the accessibility problem to the action problem: what to do with statistical data and how to connect this data with action plans? It is clearly not enough to 'merely' understand the data if it does not allow the user to transform it into (justified) decisions. An exemplar is presented, namely whether or not, having listened to the weather forecast, one should take one's umbrella, an issue that highlights intrinsic

difficulties. This analysis allows the authors to extend these questions and problems to the present-day status of the concept of 'statistical literacy' as put forward in the PISA reports, where it is clear that the ethical (and political) dimension play(s) an important role. Perhaps the most important conclusion is that the ethical and aesthetic dimensions are present throughout the whole process from the generation of the data to their societal use.

In Chapter 11, Paul Smeyers starts from the observation that statistics are everywhere but that the use of them carries with it a number of presuppositions which should not necessarily be taken for granted. One presupposition is that it is possible to represent reality. Another is the possibility of control. The attraction of statistics lies at the same time in its simplicity as well as in the belief that goes with it that it is thus possible to characterise and control reality. The chapter focuses on the question 'Why we are so eager to turn to one or other kind of statistics when trying to understand and deal with particular social practices'. Smeyers argues that philosophical problems about the structure of language (the particular metaphysical enframing (and longing) that unavoidably seems to take place), haunt us. As a result, the ideals of objectivity (bracketing the performative embeddedness) and rationality that particularly since the Enlightenment characterises our understanding of reality may be seen as emblematic of our unwillingness to live with complexity. Humans do not only long for knowledge they also seem to have an insatiable need to control. Smeyers argues that we always use concepts that invoke something general and that there is no alternative to this even if we take this generality fully into account and desperately try to avoid the victimisation of our prejudices. With the help of crime stories the chapter argues that statistics may not necessarily be the wrong road to take. Their attraction lies in the fact that they make things more simple and answer a 'human all too human need' to have some kind of grip on reality. This too is all about rhetoric and its argumentative force, something that may be easier to exploit now than ever before given the availability of super computers and web-based dissemination of what has been found to be the case.

In Chapter 12, Ian Munday discusses how statistics and the particular kind of discourse that emerges around them serve to suture the wounds in the discourse of effectiveness culture. He begins by showing how performativity operates within the British secondary school and the role statistics play there. Lyotard, who coined the term 'performativity', believed that the only resistance to 'effectiveness' was to turn to absence and silence. Following Gordon Bearn's critique of Lyotard, Munday argues that the latter's vision is hopelessly pessimistic. He therefore turns to Derrida's philosophy of language. Derrida offers a more optimistic metaphysics and his discussion of difference and iterability frames the discussion of statistics that follows. Munday begins with a discussion of numbers as particular kinds of word. Though numbers are, in a sense 'iterable' they are not iterable in a way that is comparable to other words. Indeed, in certain instances when numbers become iterable, when meaning is artificially conferred upon them, they seem to exemplify the false metaphysics of presence that Derrida's philosophy undermines: they become idealised forms of Saussure's linguistic sign which have no equivalents in ordinary language. Munday develops this train of thought in relation to statistics. Statistics

appear as surface signifiers of underlying truth. However, there is another dimension to statistics whereby they also present the promise (or threat) of absolute scepticism. This paradox serves to suture over linguistic slippages/bleeding within the discourse of performativity. It 'legitimates' the farcical 'fixing' of statistics that goes on in British schools. That said, statistics offer a limited form of power. Statistics like all numbers can never operate alone. In British and American education, the drive to boost statistics finds its voice in policy initiatives that become slogans (such as No Child Left Behind). However much slogans (that become mantras) may suture over linguistic instability/creativity this can never be wholly successful. It is within the ordinary/extraordinary operations of language that hope lies.

In Chapter 13, Richard Smith argues that the power of statistics and epidemiology, and their undoubted efficacy in many areas of the physical and human sciences, has given rise to what he calls 'metricophilia', the obsessive devotion to and faith in measurement and metrics of various kinds. He analyses three prominent texts from the year 2009: a newspaper article by Richard Layard, an advocate of measuring happiness as the standard of social progress, the report of the Stiglitz Commission, and Wilkinson and Pickett's book, *The Spirit Level: Why More Equal Societies Almost Always Do Better*. These texts, and the research on which they draw, are in different ways concerned with the failure of increased affluence to bring greater happiness or well-being to the developed world. Smith argues that the metricophilia of these texts and research projects has the effect of levelling and homogenising and requires us to think of all variations of self-esteem, happiness or well-being as essentially forms of the same thing. It leads to over-simplification and reductionism, ignoring or under-valuing crucial philosophical questions in the faith that better (and more) metrics and statistics, will tell us all that we need to know. If we need to grasp the causes of unhappiness, self-esteem and so on, it is a deeper philosophical understanding of these concepts that is needed rather than the drawing of more correlations. Achieving such understanding is an alternative to the search for the bubble referred to in the title. The texts approached in this chapter are, in different ways, coloured by rhetorical techniques and devices that are at odds with the dispassionate objectivity that statistics and science more generally have always offered.

In Chapter 14, Paul Standish argues that a sometimes neglected aspect of Thoreau's *Walden* is his examination of the various ways in which practices of accounting come into our lives and the credibility these have. This raises the question of the value and the limits of numerical accounting, including questions to do with what can and cannot be measured. It is important that Thoreau avoids both excess of faith in numbers as well as mystification of 'what cannot be measured'. Standish's discussion broaches these matters by considering the aesthetic appeal of technical languages and of numbers themselves. It offers a theoretical background in terms of Heidegger's critique of Leibniz's 'principle of reason', which in turn leads back to Thoreau. Some conclusions are drawn in terms of the ways the prominence of statistics and of numerical measures more generally cause some aspects of the world to withdraw, distorting the validity of the accounts we can give. The place of statistics in educational practice is emphasised, and a right assessment of this needs to be achieved, it is claimed, in the context of a different economy of living.

Note

1. For further information about previous work of the *Research Community*, see Smeyers (2008).

References

Depaepe, M. (1993). *Zum Wohl des Kindes? Pädologie, pädagogische Psychologie und experimentelle Pädagogik in Europa und den USA, 1890–1940*. Leuven: Leuven University Press.

Depaepe, M. (2010). Stand und Probleme der Historiographie der empirischen Erziehungswissenschaft. Metareflexionen über eine zwanzig Jahre alte Forschung als Einführung in die Tagung 'Geschichte der empirischen Pädagogik/Erziehungswissenschaft'. In C. Ritzi & U. Wiegman (Eds.), *Beobachten, Messen, Experimentieren. Beiträge zur Geschichte der empirische Pädagogik* (pp. 13–50). Bad Heilbrunn: Klinkhardt.

Gibboney, R. A. (2006). Intelligence by design: Thorndike versus Dewey. *Phi Delta Kappan, 88*(2), 170–72.

Ginzburg, C. (1989). *Clues, myths, and the historical method* (J. Tedeschi & A. C. Tedeschi, Trans.). Baltimore, MD: Johns Hopkins University Press. (Original work published in Italian in 1986).

Johanningmeier, E. V., & Richardson, Th. (2008). *Educational research, the national agenda, and educational reform. A history.* Montgomery, AL: Auburn University.

Jonçich, G. (1968). *The sane positivist. A biography of Edward Lee Thorndike.* Middletown, CT: Wesleyan University Press.

Kempe, C. H., Silverman, F. N., Steele, B. F., Droegemuller, W., & Silver, H. K. (1962). The battered child syndrome. *Journal of the American Medical Association*, 181, 17–24.

Lagemann, E. C. (2000). *An elusive science. The troubling history of education research*. Chicago: University of Chicago Press.

McCall, W. (1922). *How to measure in education?* New York: The Macmillian Company.

Porter, Th. M., & Ross, D. (Eds.). (2003). *The Cambridge history of science. Vol. 7: The modern social sciences*. Cambridge: Cambridge University Press.

Richardson, Th., & Johanningmeier, E. V. (1997). Intelligence testing. The legitimation of a meritocratic educational science. *International Journal of Educational Research, 27*(8), 699–714.

Smeyers, P. (2008). Afterword. In P. Smeyers & M. Depaepe (Eds.), *Educational research: The educationalization of social problems* (pp. 227–237). Dordrecht: Springer.

Smeyers, P., & Depaepe, M. (2003). Introduction. In P. Smeyers & M. Depaepe (Eds.), *Beyond empiricism. on criteria for educational research* (pp. 9–23). Leuven: Leuven University Press.

Smeyers, P., & Depaepe, M. (2006). Introduction. On the rhetoric of 'what works'. Contextualizing educational research and the picture of performativity. In P. Smeyers & M. Depaepe (Eds.), *Educational research: Why 'what works' doesn't work* (pp. 1–16). Dordrecht: Springer.

Tenorth, H.-E. (2010). Forschungspraxis in der Konstruktion der Erziehungswissenschaft. In C. Ritzi & U. Wiegman (Eds.), *Beobachten, Messen, Experimentieren. Beiträge zur Geschichte der empirische Pädagogik* (pp. 195–210). Bad Heilbrunn: Klinkhardt.

Tomlinson, S. (1997). Edward Lee Thorndike and John Dewey on the science of education. *Oxford Review of Education, 23,* 365–383.

Travers, R. M. W. (1983). *How research has changed American schools: A history from 1840 to the present*. Kalamazoo, MI: Mythos Press.

Wooldridge, A. (1994). *Measuring the mind. Education and psychology in England, c. 1860–c. 1990*. Cambridge: Cambridge University Press.

Chapter 2
The Lure of Statistics for Educational Researchers

David F. Labaree

> *Philosophy is written in this grand book, the universe, which stands continually open to our gaze, but the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth.*
>
> Galileo quoted in Crosby (1997, p. 240)

During the course of the 20th century, educational research yielded to the lure of Galileo's vision of a universe that could be measured in numbers. This was especially true in the United States, where quantification had long enjoyed a prominent place in public policy and professional discourse. But the process of reframing reality in countable terms began eight centuries earlier in Western Europe, where it transformed everything from navigation to painting, then arrived fully formed on the shores of the New World, where it shaped the late-blooming field of scholarship in education. Like converts everywhere, the new American quantifiers in education became more Catholic than the pope, quickly developing a zeal for measurement that outdid the astronomers and mathematicians that preceded them. The consequences for both education and educational research have been deep and devastating.

In this chapter I explore the historical and sociological elements that have made educational researchers dependent on statistics—as a mechanism to shore up their credibility, enhance their scholarly standing, and increase their influence in the realm of educational policy. I begin by tracing the roots of the urge to quantify within the mentality of measurement that arose in medieval Europe and then explore the factors that have pressured disciplines and professions over the years to incorporate the language of mathematics into their discourse. In particular, this pattern has been prominent for domains of knowledge and professional endeavor whose prestige is

D.F. Labaree (✉)
Stanford University, Stanford, CA, USA

P. Smeyers, M. Depaepe (eds.), *Educational Research: The Ethics and Aesthetics of Statistics*, Educational Research 5, DOI 10.1007/978-90-481-9873-3_2,

modest, whose credibility is questionable, whose professional boundaries are weak, and whose knowledge orientation is applied. I show that educational research as a domain—with its focus on a radically soft and thoroughly applied form of knowledge and with its low academic standing—fits these criteria to a tee. Then I examine two kinds of problems that derive from educational researchers' seduction by the quantitative turn. One is that this approach to educational questions deflects attention away from many of the most important issues in the field, which are not easily reduced to standardized quanta. Another is that by adopting this rationalized, quantified, abstracted, statist, and reductionist vision of education, education policymakers risk imposing reforms that will destroy the local practical knowledge that makes the ecology of the individual classroom function effectively. Quantification, I suggest, may be useful for the professional interests of educational researchers but it can be devastating in its consequences for school and society.

2.1 The Roots of Quantification

Alfred Crosby (1997) locates the roots of quantification in Western Europe in the 13th century. What it gradually displaced was a worldview without standardized modes of measurement, which he labels the 'venerable model.' From the perspective of this model, the world was heterogeneous, where differences were qualitative rather than quantitative and thus reality could not be reduced to common units of measurement. Fire rose and rocks fell because that was their nature, with fire returning to the sphere of fire and rocks to the sphere of earth in a four-sphere universe where air and water separated them from each other. Measuring the distance between spheres was as nonsensical as measuring the distance between God and man. Time also had a qualitative character. Years before and after the birth of Christ could hardly be measured in the same manner. And since Jesus said the day had 12 h, the length of the hour shrank as the days got shorter in the fall and then stretched in the spring. Space likewise expanded and contracted in response to spiritual importance, with maps depicting Jerusalem at the center of the world and showing east toward the top because that was the direction of Eden, toward which the world was 'oriented.' It followed naturally that the size of people in a painting was a function of their importance rather than their location in the foreground or background of the scene; saints were big no matter where they stood in the frame. Numbers were difficult to work with, since they were recorded using the first letter of the Latin word for each quantity, which meant that quantities were words and formulas were sentences.

Crosby says that quantification arose in Europe because of efforts by ordinary men to solve practical problems. Leaders and theorists were opposed to viewing the world in standardized measures, so applying the language of mathematics was a task for less elevated folk. Factors like growing trade, increased travel, and an emerging cash economy urged the process forward. As time started to become money, merchants called for reliable measures of distance, time, and accounts, which pushed sailors to develop new measures of navigation, mechanics to develop clocks, and

merchants to develop double-entry bookkeeping. They needed to keep accurate accounts, and they needed figures that could be easily manipulated, so Arabic numbers gradually made headway. But at the heart of this process, according to Crosby, was a fundamental shift in mentality toward thinking of the world in quanta. This was possible largely because of the extreme marginality and backwardness of Western Europe's intellectual and cultural life, as compared to that of the great contemporaneous civilizations of Islam, India, and China. Lacking a centralized state and an intellectual canon, people were relatively free to tinker with measurement for purely practical purposes. The quick spread of factors like cash transactions and church-tower clocks began to educate the populace in a new quantitative world in which things could be measured in fixed units.

2.2 The Adoption of Statistics by Professions

Theodore Porter (1995, 1986) picks up the story in the 19th century, exploring how professions became quantifiers. He argues that what drove the professions to adopt quantification was a growing set of challenges to their professional authority. As a technology of distance, quantification allows a professional community to make arguments that carry weight and establish validity beyond a particular time, place, and community of authorship. Its emphasis is on impersonality (Porter, 1995, p. ix). And as a technology of trust, it helps the profession gain the confidence of key actors in government, courts, and economy, who are seeking objective reasons for the choosing to follow professional advice (p. 225). The move toward quantification, he shows, was not the preferred option for most professionals. Left to their own devices, professional groups over the years have generally chosen to establish their authority through consensus within the professional community itself. But this approach only works if outsiders are willing to cede a particular area of expertise to the profession and rely on the soundness of its judgment.

Democracies in particular are suspicious of claims of elite authority, unwilling to bow to such claims as a matter of professional judgment without an apparently objective body of evidence that establishes their independent credibility. The predisposition toward objectivity that comes from numbers is a natural extension of the concept of a rule of law not men, relying on universal rules rather than personal preferences. And actors such as government bureaucrats in a democracy are especially prone to seek quantitative data to support policy actions because their own status and authority are open to question (p. 8). The United States embraced numbers early in its history for political and moral reasons as well as concerns about elite authority. The decennial census was a central mechanism for establishing the legitimacy of representative government (so that congressmen represented equal numbers of citizens), and in the early 19th century numbers became a means for assessing the state of public morality (through the accumulation of data on pauperism, intemperance, and insanity). By the 1830s, the United States experienced an explosion of the quantification of public data, with the proliferation of statistics societies and quantitative reports (Porter, 1995, pp. 195–197; Headrick, 2000, pp. 78, 87).

For our understanding of the eventual conversion of educational research to the credo of measurement, however, Porter's most salient insight is that the adoption of quantification by a profession is a function of its weakness (1995, pp. xi, 228). If a profession has sufficiently strong internal coherence and high social status, it will assert its right to make pronouncements within its domain of expertise on its own authority. To resort to supporting one's claims with numbers is to cede final authority to others. Only those professions that are lacking in inner strength and outer esteem must stoop to quantify. In particular, Porter notes that the professions and academic disciplines that are most prone to deploying numbers in support of their claims are those whose domain of knowledge is the most applied. Compared with a domain of pure knowledge, where the boundaries between its zone of expertise and the practical world are sharply defined, applied fields find themselves operating in a terrain that is thoroughly mingled with practical pursuits and thus difficult to defend as an exclusive territory (1995, p. 229). Here professionals find themselves subject to the greatest external pressures and the strongest need to demonstrate the credibility of the claims through quantitative means. Such is the terrain of educational research.

2.3 Educational Research as a Soft and Applied Field

American educational researchers in the early 20th century took the plunge into quantification. This was the era of Edward L. Thorndike and Lewis Terman; of the proliferation of intelligence tests and other standardized assessments in schools; and of the development of scientific curriculum, which built on testing to track students into suitable studies. It was the period chronicled by Stephen Jay Gould (1981) in *The Mismeasure of Man* and by Nicholas Lemann (2000) in *The Big Test*. It was when psychology became the dominant discipline within education by embracing quantification more quickly and wholeheartedly than other domains in the field. A useful way to figure out what this happened is to apply to the field of educational research Tony Becher's (1989) analysis of the link between knowledge domains and the social organization of disciplines. In doing so I will draw on my own elaboration of this application to education developed elsewhere (see Labaree, 1998a, 2004, chapter 4).

Becher begins by locating the knowledge domain of individual disciplines on a scale between hard and soft knowledge and between pure and applied knowledge. Then he proceeds to explore the consequences of this knowledge location for the organization of research endeavors within each discipline. In this analytical scheme, educational research is classified as very soft and very applied; and the consequences of this are devastating for the ability of educational researchers to accumulate knowledge, defend it from outsiders, develop a coherent account of the field, build on previous work, and convince policymakers to take their findings seriously.

The difference between hard and soft fields of study is familiarly understood in terms of distinctions like quantitative and qualitative, objective and subjective, and definitive and interpretive approaches. Of course, the hard–soft distinction is