HARVEY GOLDSTEIN

Harvey Goldstein

30 October 1939 – 9 April 2020

elected Fellow of the British Academy 1996

by

JOHN GRAY
Fellow of the Academy

IAN DIAMOND
Fellow of the Academy

FIONA STEELE
Fellow of the Academy

Harvey Goldstein died in April 2020 aged 80, an early victim of the Coronavirus pandemic. During the course of his career he made getting on for 400 published contributions to the development of statistical methods and to the substantive analysis of various educational, social and medical issues. He is particularly noted for his pioneering contributions to the development and dissemination of multilevel modelling and for his sustained critiques of so-called league tables of institutional performance.

HARVEY GOLDSTEIN

Professional life and contributions

In 1977 the London Institute of Education (IoE) advertised a Chair of Statistical Methods and appointed Harvey Goldstein to it. At first glance this was a routine appointment—several professorial positions were being created across the Institute and a new head of department was needed for the Department of Measurement, Analysis and Computing. However, it was an unusual appointment in at least two respects. First, Harvey had been a lecturer at the London Institute of Child Health and, whilst he had already published quite extensively, relatively little of his output had had much to say about educational matters. And second, Educational Research was still in the process of establishing itself as an area of enquiry with its own distinctive approaches. Some of the major US graduate schools of education had a dedicated chair in statistics but there had never previously been such a chair in the UK.

Expectations were probably quite low as well. Interest in statistical developments was largely confined to those with backgrounds in psychology, and most researchers who inclined towards quantitative methods saw statistical analysis as a relatively straightforward matter. The department was one of the smallest at the Institute and was perceived, at that time, as performing mostly a service role. Furthermore, there had been a major swing towards qualitative approaches during the 1970s and a prevailing scepticism about the potential of quantitative methods. In short, on one reading of the overall climate for this new chair, the prognosis was not good.

On the other hand, low expectations offered freedom to pursue issues as Harvey saw fit, and the field was calling out for leadership. At the age of 38, a good thirty years of contributions were in prospect. And, within a short time of his appointment, it had become apparent that there was a new and significant player on the block—someone who was not afraid of challenging authority but, equally, one who was committed to ensuring that the highest standards of statistical analysis were put at the disposal of the nation's policy-makers, researchers and practitioners.

Early contributions

Harvey's interest in statistics was fostered at the University of Manchester where he took optional statistics courses, in addition to his degree in mathematics; these were given by the politically radical statistician Toby Lewis, who became a lifelong friend and mentor. He undertook an undergraduate survey into housing in the London area of Notting Hill which helped him to understand that statistics could improve people's lives. On graduation he proceeded to the Diploma in Statistics at University College London where he also worked as a research assistant for two years (Mortimore 2020). Few publication lists kick off with a short contribution to *Nature*, albeit one jointly

authored with Sir Richard Doll (and other co-workers) who were busy establishing the links between smoking and cancer (Jacobs et al. 1963).

It was a period where statisticians often worked as research assistants before embarking on careers as lecturers; Harvey moved to the Institute of Child Health as a Lecturer in 1964. Here he worked with Jim Tanner on a number of studies associated with child development. He spent some time developing growth curves but found them somewhat frustrating and a largely theoretical exercise. For example in 1971 he wrote: 'It is very difficult to find any successful contributions to biological knowledge that have resulted from the study of theoretical growth curves', whilst also noting: 'In the past, I have myself spent time pursuing this new philosopher's stone. I never created gold, and nor to my knowledge has anyone else' (Goldstein 1971). This paper also provides an interesting historical aside when he notes that some progress might be made using computer programmes (although he doubted it!).

Tanner was the visionary research leader who, in 1946, had started the first national birth cohort in the UK, the National Survey of Health and Development. This was to become the forerunner of the suite of UK birth cohorts which represent a jewel in the UK's social and medical science infrastructure. Harvey also met and started a long collaboration with Neville Butler, another far-sighted researcher, who masterminded similar cohorts for 1958 and 1970. The report *From Birth to Seven* was one influential outcome of their partnership, bringing together a range of important evidence on educational, psychological and health-related outcomes (Davie et al. 1972) that attracted widespread interest in policy circles.

In 1969 he published a superb article which demonstrated the power of longitudinal data and the need for careful analysis (Goldstein 1969); and soon after, with Butler, published a seminal piece of longitudinal analysis which demonstrated the negative effects of smoking during pregnancy on birthweight and perinatal mortality (Butler and Goldstein 1972). A social habit with relatively little impact on the foetuses of advantaged mothers could prove a serious risk for those born in poverty.

In subsequent years this result was the subject of some debate with the nub of many arguments being that it was not smoking per se that caused low birthweight but rather that women who smoked had different characteristics and the smoking was a proxy for such characteristics. Harvey considered these arguments in great detail (Goldstein 1977) and, in the first of a number of articles over the years which addressed criticisms of his work, concluded: 'The main objections to the statistical case for a causal relationship between smoking in pregnancy and birthweight and mortality do not bear much scrutiny.' Around this time, still working on the 1958 cohort, he published another seminal paper with Michael Healy on bone maturity (Healy and Goldstein 1976).

Harvey himself was responsible for planning much of the strategy for the next stage of the 1958 birth cohort whilst working at the National Children's Bureau (1972–7). He also became ensconced in the potential for causal inference from these data—as long as the analysis was careful, took into account non-ignorable strata, and was focused on a clear problem.

At the same time his influence as an international statistician started to grow. He was one of a group of experts, organised by the World Health Organization and including Neville Butler, who were invited in 1970 to advise the post-revolutionary government of Cuba about population statistics and child development. Subsequently he returned on a number of occasions in the early 1970s to give statistical advice and support the development of a Cuban birth cohort study.

The move into education

In 1977 Harvey moved to the IoE. Here, not surprisingly, he directed most of his interests towards educational research, most notably, in due course, to issues related to school effectiveness and educational assessment, areas where quantitative traditions were already established. However, he still retained an interest in longitudinal studies, publishing a textbook on their design and analysis (Goldstein 1979). He also made a plaintive plea in 1981 for statistics in educational research to be taken more seriously. He bemoaned the fact that in the UK few statisticians were interested in educational research and that quantitative research in education had been largely restricted to psychometrics (Goldstein 1981). He was careful not to criticise psychometrics as such but, rather, pointed out the many important non-psychometric problems that remained to be addressed.

By comparison with his later prolific contributions on educational policy issues, his early excursions were comparatively modest. A large, edited volume of reviews and critiques of all the various tests of educational matters was one contribution (Levy and Goldstein 1984). He also became interested in the use of statistical modelling techniques in the construction and analysis of educational tests. He was particularly concerned about the so-called Rasch model. His criticisms of the use of this approach at the National Foundation for Educational Research (at that time probably the major producer of educational tests in the UK) led to a rather public falling out. There were also contributions on the comparability of standards at different exam boards, again a rather 'hot' topic. Finally, there was an extensive evaluation of the Assessment of Performance Unit, one of the forerunners of subsequent moves to hold all schools more publicly accountable for their pupils' performance (Gipps and Goldstein 1983). The Unit was abolished when the 1988 Education Reform Act extended testing to all

pupils. Its influence lived on, however; many years later he was to return to its approach as a possible model for dealing with some of the vexed issues of accountability which plagued debates about so-called league tables.

The 'unit of analysis' question

Every generation of researchers encounters a problem or set of concerns that are so central to the development of the discipline that their solution becomes a priority. The 'multilevel' problem falls into this category—the lack of a consensual view from statisticians caused confusion for more than a decade. When results were computed in two 'equally respectable' but different ways, researchers were finding that they got different answers. Sometimes the differences were too small to worry about but, on occasion, they were sufficiently discrepant to be worrying. It was to take several more years before some kind of solution was to emerge.

The origins of the multilevel approach are to be found in the so-called unit of analysis question. Some appreciation of the issues began to emerge in the late 1960s; by the 1970s the issue had become more pressing. It clearly affected the analysis and interpretation of many important studies of educational issues at that time. It was central to the Headstart Planned Variation Study, which was an attempt to rescue pre-schooling from a disastrous (negative and premature but nonetheless highly influential) evaluation. It figured prominently elsewhere as well: in studies of school and teacher effectiveness for example, of the success or otherwise of comprehensive school reforms, and of attempts to judge the performance of local education authorities.

Initial steps towards multilevel modelling (MLM)

It was about this time that Harvey began to write about what would become, perhaps, his greatest contribution, namely the analysis of hierarchical (or multilevel) data. In an unpublished paper in 1983 he wrote down the general form of a multilevel model for the first time and suggested it could be estimated by iterative generalised least squares. He was not doing this for analytic curiosity but because he had been increasingly concerned about proper ways to compare schools' performances. This is explained in an article on the methodology of school comparisons (Goldstein 1984). He starts by pointing out that schools differ in the academic standards of their intake; and that there can be large heterogeneity between schools of the same broad type. Hence it is critical to build a measure of 'value added' into any comparisons between schools.

He then went on to discuss units of analysis, basing his argument on the observation that education is usually hierarchically organised with *pupils* studying in *classes*

which are part of *schools* which are, in turn, in *local education authorities*. Thus, as he points out, at any one level, all the units at that level share the same set of characteristics of the unit at the next higher level within which they are grouped; and so: 'A class can be classified by the average test scores of the pupils in it, or, for example, by the standard deviation of those test scores.' He concludes that, while school or local authority level analyses have some use, their contributions are limited and that a full multilevel design is necessary to make proper and fair comparisons.

Murray Aitkin and Nicholas Longford were also working on these issues at Lancaster, where they were acting as consultants to the SSRC-funded Contexts project, led by John Gray, into 'the use and interpretation of exam results as measures of school performance' (Aitkin and Longford 1986; Gray et al. 1986). Harvey and Nicholas Longford first presented their ideas at a small seminar in Edinburgh where a member of the audience pointed out that they were both tackling essentially the same problem, albeit in rather different ways.

A few months later they met again. John Gray takes up the story:

> A number of researchers were gathered in Lancaster to hear Murray Aitkin's re-analysis of the controversial research on *Teaching Styles and Pupil Progress* (Bennett 1976). It had been a long day and it was with no great enthusiasm that I attended an unscheduled after-dinner seminar. Aitkin made a few introductory remarks before his co-worker Longford picked up the chalk and attacked the blackboard, covering it with hieroglyphics. After about 15 minutes he stopped and handed over to Harvey Goldstein. Goldstein worked with more deliberate speed; his demeanour suggested a more elegant solution. After a few brief exchanges the two of them shook hands and sat down. The person sitting next to me said it had been 'an historic occasion'. It took me a couple of years to understand why history had been made; it was to take me the better part of the next decade to become fully familiar with the power of 'multilevel' modes of thought. (Gray 1998: 8)

It turned out that the educational world was full of so-called nested structures but at last a solution to the 'units of analysis' question was in prospect—and furthermore a potentially exciting one. Over the next three years Harvey threw himself into creating the conditions for multilevel modelling to take root. A paper outlining the statistical theory of iterative least squares estimation methods was prepared for publication in *Biometrika* (Goldstein 1986). A year later the first edition of his book *Multilevel Models in Educational and Social Research* was published (Goldstein 1987). At the same time ambitious proposals for further developments were put to the ESRC research council.

Establishing the Centre for Multilevel Modelling

The hard work paid off and a sizeable research grant was forthcoming which underpinned the work of the newly constituted Multilevel Modelling Project. Several factors contributed to its longer-term success. In Harvey multilevel modelling had a 'champion' who was prepared to put in the hours to turn theoretical ideas into practical applications. He appreciated that a substantial research and training programme lay ahead which needed to be informed not only by sophisticated statistical approaches but high-quality training courses. In the process he built up a small team of individuals who were equally committed to the overall endeavour, including Jon Rasbash, Min Yang, Fiona Steele and Bill Browne. Educational research also turned out to be fruitful ground for exploring 'nested' structures, and the team spent many hours engaging with researchers at all levels of expertise, and not just the most advanced.

Some contextual factors were also at play. The Institute of Education was at the centre of an influential worldwide network so there was a steady stream of potential new recruits. Previous links with past collaborators in the medical and health sciences were maintained, whilst London had numerous social scientists, many of whom, they discovered, were also exploring multilevel structures.

Luck probably came into it as well. Competition with other researchers was not particularly intense. Aitkin and Longford seemed less interested in disseminating their 'Variance Components' approach to educational researchers and, some years later, Aitkin returned to his native Australia whilst Longford (1988) went off to work at the Educational Testing Service in Princeton. In the USA, meanwhile, Bryk and Raudenbush (1992) at Chicago were busily promoting their Hierarchical Linear Modelling (HLM) approach but made only occasional forays into Europe. Their network accounted for much of North America but there was room for more than one approach, and Multilevel Modelling, based in the London Institute's research powerhouse, became the method of choice across many other parts of the world.

Harvey embarked on another period of intense activity and the benefits of his incredibly productive collaboration with the computer programmer, the late Jon Rasbash, became obvious. He foresaw the need, if multilevel modelling was to become used widely, for easily accessible software, and contributed to the development of a series of software packages (starting with ML2 in the late 1980s, and followed by ML3, MLn and MLwiN) which offered the prospect of considerably more than the two- and three-level structures that had prevailed up till then. These packages became very widely used across many of the social sciences and beyond. The team also put considerable effort into running workshops to enable scholars to learn how to use the approach and resources; if users stumbled in grasping the complexity, the team was usually there to help them.

In addition he worked extensively with leading statisticians engaged in the Economic and Social Research Council's programme 'The Analysis of Large and Complex Datasets'. This featured teams from all the UK's major social statistics centres. It is interesting that Harvey's work was similar in some respects to research going on at Southampton where Tim Holt, Chris Skinner and Fred Smith were working on the analysis of complex survey designs;[1] in a cluster sample there would be potentially correlations between the observations in a cluster, and thus it was that the power of multilevel modelling in many different arenas was explored. Again, the opportunities opened up for initial two- and three-level designs (pupils within classes or pupils within classes within schools) were, in due course, superseded by increasingly complex possibilities.

The wider impact of MLM on research

During that first decade the research team had begun to secure its place and the project had become a Centre at IoE with an increasingly international reputation and clientele. Multilevel modelling had circled the world and been applied in many disciplines beyond education. Problems which could once only be explored by booking time, months ahead, on the world's fastest computers, could now be solved on desktop machines. And importantly, whilst MLM did not necessarily produce new answers, it sometimes did.

In retrospect, if the multilevel rationale was accepted, a range of previous studies needed to be re-examined in some detail to test their robustness under these different and more stringent conditions. In practice such re-examinations rarely occurred—the implications might have been too disturbing to the research 'inheritance'.

Harvey made major contributions to several vexing educational questions. One such example was work on the effects of smaller classes on pupil performance. This has long been a central question for educational researchers but also one which has been dogged by poor research designs, poor statistical analyses and, on occasion, counter-intuitive conclusions. In the late 1990s he turned his attention to this problem in a fruitful collaboration with Peter Blatchford. The results of their research showed that class size did matter but especially so early on in a child's career; as children progressed through primary school the effects got less significant (Blatchford et al. 2003). Harvey also undertook a reanalysis of the influential STAR study (an early example of an RCT in education) subjecting it to a more rigorous analysis employing a multilevel approach (Goldstein and Blatchford 1998). This secondary reanalysis did not

---

[1] For Chris Skinner, see Ray Chambers, Ian Diamond, Tim Holt, Paul Smith and Fiona Steele, 'Chris Skinner, 1953–2020', *Biographical Memoirs of Fellows of the British Academy*, XIX (2020), 377–93.

contradict the earlier conclusions but did support the emerging wisdom that class size effects were at their most potent during the early years.

The perspectives ushered in by multilevel modelling went on to have considerable influence on the design of subsequent research. Sample sizes tended to become larger, and considerably more attention was paid to ensuring that there were adequate numbers of higher level units in research studies, which typically meant involving more schools than had traditionally been deemed necessary (see for example, the research design reported in Mortimore et al. 1988 on the relative effectiveness of primary schools). Looking back, the need for adequate numbers of higher-level units seems self-evident but, at the time, the consequences seemed rather challenging.

On the (severe) limitations of league tables

By the early 1990s schools' exam results were firmly in the public domain. There were numerous criticisms of the limitations of the resultant 'league tables' produced and published by the national press. Harvey had by this point been working with a number of different school effectiveness researchers whose work touched directly on issues of school quality and public accountability; colleagues working on these and related issues at IoE included Peter Mortimore, Pam Sammons and Sally Thomas. For a while the incorporation of 'value-added' estimates was seen as some kind of antidote to the simplistic measures being employed. School effectiveness research more generally, and multilevel modelling in particular, appeared to offer a way forward.

There was a major problem, however, which Harvey was quick to point out and indeed continued to draw attention to for the next two decades. Using multilevel models, a confidence (or 'uncertainty') interval for each school's performance can be estimated. These are helpful for drawing appropriately circumspect judgements about individual schools for research purposes, but dynamite as far as crude attempts to rank schools along the lines of football league tables were concerned. He was able to demonstrate that, even if you believed the measurement of school performance was valid, it was extremely likely that there would be no statistically measurable differences between most schools. The width of the confidence intervals for each individual school was such that it was not possible to distinguish between their performances. In most samples examined, just a handful of schools might be performing considerably better than predicted and another handful considerably worse.

There were also other kinds of unpredictability. Schools' value-added performances, for example, were not that stable over time—knowing the results for one year did not necessarily give a very good purchase on next year's results or further down the line (Gray et al. 1995). This point undermines quite fundamentally the government's

simultaneous promotion of school value-added performance for parental school choice purposes, since such inferences implicitly involve extrapolating the current performance of schools some five years or more into the future when the children of parents choosing take their end of school exams (Leckie and Goldstein 2009).

In truth, multilevel analyses showed that most schools most of the time produced the kind of pupil progress you would predict from knowledge of their starting points. If you were of a 'league table' mindset you would basically have just three groups of schools; anything more finely ranked could not be justified. A paper with David Spiegelhalter showed that, while it is possible to make comparisons between institutions, great care should be undertaken when doing so regardless of whether the field is education or health (Goldstein and Spiegelhalter 1996).

The paper was written not long after the Department for Education conceded the need for value-added tables, but in this paper Harvey noted that while value-added tables are an improvement on raw scores they still suffer from many other limitations. Indeed, a paper two decades later was to express his continuing concerns about their (severe) limitations (Leckie and Goldstein 2017).

Notwithstanding what might seem like fundamental criticisms, Goldstein and Spiegelhalter concluded that comparing institutions can be a valid exercise. However, this should be undertaken in a spirit of collaboration rather than competition. Although not mentioned in the paper, work that Harvey did with the Hampshire Local Education Authority (Goldstein et al. 2000) pointed in that direction. He calculated multilevel rankings but subsequently schools were just given their general position and the reasons for it, rather than seeing the league tables themselves. He argued that this enabled schools to work on what needed to be improved without being in the full gaze of potential public criticism. In a lecture given shortly before his death he summed up his views: 'I am not advocating that we drop the idea of publicly accountable systems, rather that we move away from naïve and misleading presentations of evidence towards a more rational approach' (Goldstein, 2020).

The previous two decades had been highly productive and recognition from his academic peers followed. In 1996 he was elected to the Fellowship of the British Academy, the first instance ever of a researcher based in an education department receiving this honour. Two years later he was awarded the Guy Medal in Silver from fellow statisticians at the Royal Statistical Society. The citation drew attention to his 'work in developing and applying multilevel models in educational and social research, making a new set of important analysis techniques available to a wide range of fields of application'. MLwiN, his main statistical legacy, is still being actively maintained and over its history has had around 30,000 users worldwide. The citation also made particular reference to the paper with Spiegelhalter.

The move to Bristol

With Harvey's formal retirement from the Institute looming, the multilevel team began to think about its longer-term future. A decision was made to move the Centre to Bristol University with Jon Rasbash taking over as Director and Bill Browne and Fiona Steele also making the move. In 2005, aged 65, Harvey followed, taking up a part-time professorship at the School of Education whilst continuing to live in London. Far from 'retiring', however, this post-retirement move seems to have inspired another bout of energy which he largely sustained until his death.

Bristol proved a fruitful location with its strong groups of educational researchers and social scientists. Harvey also created links with the ALSPAC team in the School of Social and Community Medicine (which was natural given his longstanding interests in birth cohort studies). The Centre received significant funding from the ESRC as a node of the National Centre for Research Methods which continued for nine years with Harvey as a co-investigator. Following Jon Rasbash's untimely death, he continued working with Bill Browne and Chris Charlton at Bristol on software developments, supporting a large and increasingly worldwide user community. By the time of his death, his 'textbook' on multilevel modelling had clocked up some 9,000 mentions on Google Scholar.

Alongside his role at Bristol, Harvey took up a part-time appointment at the Great Ormond Street Institute of Child Health, reconnecting with his earlier interests and stamping ground in medical statistics. He also took up visiting professorships elsewhere, including the University of East Anglia, the London School of Hygiene and Tropical Medicine, the Institute of Fiscal Studies and the Australian Catholic University in Brisbane.

Until the late 1990s, Harvey's methodological work had used traditional frequentist approaches. He then developed an interest in Bayesian estimation via Markov chain Monte Carlo (MCMC) methods, working with Bill Browne. The main attraction of MCMC methods was that their flexibility allowed Harvey to move closer to his goal of estimating realistically complex statistical models. An early use of MCMC was to provide a computationally efficient way of estimating multilevel models for non-hierarchical structures (Browne et al. 2001), including cross-classified models to estimate effects of primary and secondary schools or school and neighbourhood effects on educational attainment, and multiple-membership models to allow for pupil mobility between schools over time (Goldstein et al. 2007).

Another attraction of MCMC methods was their facility for handling categorical responses, which found applications in the analysis of event history data (Steele et al. 2004) and multivariate models for mixtures of categorical and continuous responses (Goldstein et al. 2009). His work with James Carpenter and Mike Kenward on a

general class of multilevel models for multivariate data formed the basis of important and influential research on multiple imputation for missing data (Goldstein et al. 2014). Previous work on multiple imputation had largely focused on continuous variables and had ignored complex data structures, which was inadequate for most real-world applications.

Harvey also maintained an interest in models for data where the covariates are subject to measurement or misclassification error (Goldstein et al. 2008). Combined with his research on multiple imputation, this work had particular relevance for probabilistic data linkage which Harvey, with Katie Harron and colleagues at ICH, framed as a missing data problem (Goldstein et al. 2012; Goldstein and Harron 2015). In recent years, the linkage of different administrative datasets, particularly in health, and of survey and administrative data, has been increasingly widespread, and the work of Harvey and co-authors has been making a substantial impact on the field by improving the robustness of the information made available to researchers.

At the time of his death he was working on and applying models for generating synthetic data-sets in studies where sensitive data could not otherwise be shared openly.

Contributions to longitudinal studies

Harvey's commitment to longitudinal studies spanned his career. However, over time, he became increasingly aware of the extent to which they were vulnerable to various threats including short-term funding and the whims and fashions of policy-making. John Bynner, a longstanding friend and, from time to time, colleague and himself an influential advocate of longitudinal approaches, takes up the story:

> Harvey was a system builder. Back in the mid-1990s he became convinced that collaboration between the large-scale UK longitudinal studies was crucial to their survival. As protection he was instrumental with others in setting up first the London-based Joint Centre for Longitudinal Studies to be followed in 2005 by the wider ranging thinktank Longview with the aim of promoting longitudinal research to its three main constituencies—policy-makers, scientists and the general public. The Longview approach, inspired by Harvey, comprised a productive mix of expert groups, surveys and specialists from the UK and overseas to debate findings and draw conclusions… This work set the scene for the new era that Harvey had fought for in which longitudinal and life course studies were coming to be seen as national and international resources. (Bynner, 2020)

Bynner has also drawn attention to Harvey's methodological contributions which have improved the rigour of longitudinal analyses: the comparative ease with which multilevel structures can now be explored in longitudinal datasets has helped to usher

in a new series of insights, whilst the work on sample attrition problems and missing data, problems which handicap most longitudinal work, can now be dealt with somewhat more robustly.

His various interests in this field of enquiry came together in due course with the launch of a new journal *Longitudinal and Life Course Studies* in 2009. As Heather Joshi (an editor of the journal and herself a leading player in the development of the wide-ranging Millennium Cohort Study) has noted, he was not only a section editor for the journal but also a 'prolific' author. 'He appeared eight times in the first ten volumes of the journal—articles, research notes and contributions to debates—on topics relating to statistical methodology, survey design and administrative record linkage in the context of longitudinal data collection and analysis' (Joshi 2020). He was still battling for longitudinal studies in the later stages of his career. Joshi reports that he played 'a leading and creative role in the ambitious but ultimately unsuccessful attempt to establish another (interdisciplinary) national cohort study in 2015' (Dezateux et al. 2013).

The man behind the numbers

Harvey is survived by his wife Barbara (née Collinge), whom he married in 1970, and son Tom. He enjoyed listening to classical music and playing the flute in a wind band and orchestra. Another interest was cycling and he and his wife Barbara produced a book of rides *Wheel Around Norfolk: Cycle Rides for All* (1994).

Political beliefs and values

Born in Whitechapel in the East End of London, Harvey was subsequently brought up in Edmonton. His father Jacob used evening classes to become a semi-skilled engineer; his mother Millie (née Belanoff) was a hat-maker. When he was just five his mother died of a heart attack. Harvey was then brought up by his Jewish grandparents, living with them until his father returned from military service. He attended Hendon County Grammar school (now Hendon School) and excelled in maths and science before reading pure maths at Manchester University (Mortimore 2020).

Although as a child he had always been aware of his father selling the *Morning Star* on the streets of North London, his political education truly began when he reached 12 and his father remarried (at this point he acquired a stepsister). Both his parents were ardent communists and as soon as he was old enough (15) he joined the Young Communist League. As with many such families with communist sympathies

(and Eastern European backgrounds), following the Soviet invasion of Hungary in 1956 when Harvey was age just 17, he resigned from the party (Mortimore 2020).

Harvey had no formal religious affiliation and his parents made no attempt to give him one. He was not a practising Jew, and certainly not a Zionist, but felt his Jewishness was part of his identity. He was a periodic (and usually published) letter writer to the *Guardian*, but was very surprised indeed when they rejected every one of his more recent letters about anti-Semitism, stating his opinion that there was insufficient evidence of anti-Semitism against Corbyn and the Labour Party.

After giving up on communism, his next political step was switching to socialism via support for the Labour Party. Being often critical of the party leadership, who he believed had never really taken the party in a proper socialist direction, he finally moved on, having a particular disdain for New Labour under the Blair–Brown regimes and favouring the Green Party instead.

But these judgements were tempered by the behaviour and politics of his local party. Compared with past MPs, the current Labour MP since 2015, Catherine West, with international credentials and an EU supporter, was well regarded and a notable exception. However, this appeal was not sufficient to shift his vote away from Green to Labour in the general elections which followed. His reasoning was driven by the geological significance of the climate change crisis. He was, amongst other groups, a supporter of Scientists for Global Responsibility, the Socialist Education Association and the Radical Statistics Group.

Throughout his career Harvey was involved in a variety of relationships and consultancies with outside bodies: government departments who sought his advice such as the Department for Education in its various guises over the years, the Home Office and the Office for National Statistics; and learned societies including the Royal Statistical Society, the British Academy and the British Educational Research Association. He was a very active member of the British Academy's Section S4 which covers Sociology, Demography and Social Statistics as well as a number of other areas including Education and, until recently, Management and Business Studies; he was also a founder member of the Education Group which was afforded full section status at the Academy's AGM in July 2020. He played an important role in the Academy's public impact programme, producing a well-received pamphlet for them on league tables.[2] In addition he participated in a range of bodies concerned with higher education including the HEFCE Research Assessment panel for the

---

[2] This short guide for head teachers and governors (Goldstein 2013) followed a British Academy report he had co-authored on the use of league tables in the public sector more generally (Foley and Goldstein 2012). A decade before that, he had co-edited for the Academy a short volume on educational standards (Goldstein and Heath 2000).

2001 exercise; and made various contributions to the work of the Economic and Social Research Council.

Nearing formal retirement he was eventually put forward for an honour. One of the later versions of his CV records, very unusually, in the section on Awards and Honours: '2006: *Declined* CBE in the New Year Honours List'. He was pleased to see the achievements of others recognised in this way but to have accepted one himself would have been to ignore what he saw over many years as the continual rejection by government ministers of his advice; he probably felt to have accepted an honour might have severely compromised his own much-valued independence. An honorary doctorate from the Open University in 2001 was more in line with his principles.

Teacher, mentor and colleague

In many US graduate schools some grounding in statistics is compulsory. This has tended not to be the case in the UK. Consequently Harvey had relatively few registered students at the Institute—nonetheless, he influenced many more. Much of his teaching was conducted in the demanding circumstances of the public domain and covered a very diverse range of groups. Many of those who encountered him describe the experience in ways that are typically reserved for 'inspirational teachers' and mentors.

Jo-Anne Baird, now a professor at Oxford, describes one such encounter when, as a young researcher, she was expected to present work on exam comparability to a high-level panel Harvey was sitting on: 'To say the least, we all found this a daunting task. Harvey was *the* expert on multilevel modelling and was a harsh critic; however, he liked my work and took an interest, which was very encouraging for a young researcher new to the field ... I took his advice … I will always remember Harvey as intrepid, independent and principled. It could be difficult being on the receiving end of his critique but he was ultimately interested in social justice …'

Gemma Moss, now a professor at the UCL Institute of Education, writes: 'In person I will remember him most for his sense of humour, the breadth of his interests and his conversation. As an academic I [was] in awe of his approach to taking on difficult subjects in education, particularly when injustice seemed to be being done, and the forensic quality of the research he brought to bear on those topics. Finding a way to be engaged and political on subjects that matter, whilst upholding the very highest standards of academic rigour in research, is a rare combination.'

Lindsay Paterson, a professor at Edinburgh, describes one such public occasion of Harvey living by the values he espoused. He recalls a meeting at which 'a very senior academic urged caution in being too critical of the government's enthusiasm for league tables, lest we lose all influence on how they might be developed. Harvey,

equally senior, answered that point by asserting his steadfast principle that we, as scientists, ought to always follow the evidence before we follow ideology or pragmatism.'

Peter Blatchford, a professor at IoE, writes: 'Harvey was one of the most brilliant people I have had the privilege to work with. He had a rare facility to be technically adept while also quick to see the big picture.'

Lindsay Paterson also writes about Harvey's ability to communicate to a wide range of different audiences: 'What really sticks in my mind was the clarity and patience of his explanations to the very diverse group of training course members, all of whom were in occupations like school teaching, educational advisory work and government. There was never any hint of pomposity or impatience: he always took each student from where they were …'

George Leckie, a former research student, long-term collaborator and now a professor at Bristol, picks up another side of the teaching relationship. He writes: 'Harvey and I were some 40 years apart in age and we were perhaps an unlikely pairing but he would go on to mentor and support me throughout the next 15 years ... He often presented his ideas in the form of dense equations with only the occasional string of words here or there as offer of explanation. I had the distinct feeling that he deliberately presented his ideas to me in this way to make me really learn my trade.'

Talking of his strengths as a supportive colleague Kathy Sylva, now an emeritus professor at Oxford, refers to a period in the early 1990s when female professors at IoE were 'treated courteously as a different species'. Consequently 'the women began to meet separately … to support one another in what was a masculine world. No-one who knows Harvey will be surprised that we regularly invited him to our "female-only drinks".' 'Harvey was his own man, not defined by gender or discipline, and certainly not one of the "male club" which ran the Institute at that time. He made us feel welcome at the Institute, laughed with us at antiquated conventions and advised us in our uphill campaign for "equal pay for equal work".'

Living with and by the evidence

A well-attended conference was organised at the Institute of Child Health to celebrate Harvey's 80th birthday. During the proceedings he delivered a prestigious Otto Wolff Lecture. He reviewed many of the themes which had animated his career, focusing on how statistics are viewed and used in the public arena (Goldstein 2020). Where is all this heading, he asked, and where does this leave us?

Answering his own questions, he summarised the values which had, he hoped, informed his own behaviour and career: 'I have little doubt that, ultimately, real evidence can win out if the issue is serious enough … The important thing for researchers

is not to give up. The research and the publicising of the implications of that research, along with public critiques of evidence of abuse or suppression, need to continue. All of this is difficult but I think there is an ethical imperative to try to do it.'

The nub of his advice was that, if one wants to influence public policy on potentially contentious issues, one is in for the long haul. From his own work the debates about smoking, people's attitudes towards it and its often disastrous consequences spring to mind. So too do the seemingly interminable debates about the manifold weaknesses of various performance indicators and 'league tables'. In due course and in both cases, significant changes have been ushered in but the time-scales have been measured in decades rather than years. Harvey had the personal stamina and resilience to battle on for what he believed was right. For him, statistics, rigorously conducted and with a clear eye on their limitations, were a uniquely important way of speaking truth to power.

References

Aitkin, M. & Longford, N. (1986), 'Statistical modelling in school effectiveness studies (with discussion)', *Journal of the Royal Statistical Society, Series A*, 149(1): 1–43.

Blatchford, P., Bassett, P., Goldstein, H. & Martin, C. (2003), 'Are class size differences related to pupils' educational progress and classroom processes? Findings from the Institute of Education class size study of children aged 5–7 years', *British Educational Research Journal*, 29(5): 709–30.

Browne, W., Goldstein, H. & Rasbash, J. (2001), 'Multiple membership multiple classification (MMMC) models', *Statistical Modelling*, 1(2): 103–24.

Bryk, A. & Raudenbush, S. (1992), *Hierarchical Linear Models* (Newbury Park, CA: Sage).

Butler, N. & Goldstein, H. (1972), 'Cigarette smoking in pregnancy: Its influence on birthweight and perinatal mortality', *British Medical Journal*, 2(5806): 127–30.

Bynner, J. (2020), 'Obituary: Harvey Goldstein', *Society for Longitudinal and Life Course Studies Newsletter*, 37: 4.

Davie, R., Butler, N. & Goldstein, H. (1972), *From Birth to Seven* (London: Longman).

Dezateux, C. & 16 others (2013), 'Life Study: a UK-wide birth cohort study of environment, development, health and wellbeing', *The Lancet* (29 November), 382: 531.

Foley, B. & Goldstein, H. (2012), *Measuring Success: League Tables in the Public Sector* (London: British Academy).

Gipps, C. & Goldstein, H. (1983), *Monitoring Children: an Evaluation of the Assessment of Performance Unit* (London: Heinemann).

Goldstein, H. (1969), 'Longitudinal studies and the measurement of change', *The Statistician*, 18(2): 93–117.

Goldstein, H. (1971), 'The mathematics background to the analysis of growth curves', *Proceedings International Congress in Pediatrics, 29 August—4 September 1971, Vienna, Austria*, XV, 5–10.

Goldstein, H. (1977), 'Smoking in pregnancy: some notes on the statistical controversy', *British Journal of Preventive and Social Medicine*, 31(1): 13–17.

Goldstein, H. (1979), *The Design and Analysis of Longitudinal Studies* (London: Academic Press).

Goldstein, H. (1981), 'Statistics without statisticians: the case of education', paper delivered to the RSS Social Statistics Conference, April.

Goldstein, H. (1984), 'The methodology of school comparisons', *Oxford Review of Education*, 10(1): 69–74.

Goldstein, H. (1986), 'Multilevel mixed linear model analysis using iterative generalised least squares', *Biometrika*, 73: 43–56.

Goldstein, H. (1987), *Multilevel Models in Educational and Social Research* (London: Griffin; New York: Oxford University Press) (first of four editions).

Goldstein, H. (2013), *School League Tables: a Short Guide for Head Teachers and Governors* (London: British Academy).

Goldstein, H. (2020), 'Living by the Evidence', The 2019 Otto Wolff Lecture, *Significance* (Royal Statistical Society), 17: 1.

Goldstein, H. & Spiegelhalter, D. (1996), 'League tables and their limitations: statistical issues in comparisons of institutional performance', *Journal of Royal Statistical Society, Series A*, 159(3): 385–409.

Goldstein, H. & Blatchford, P. (1998), 'Class size and educational achievement: a review of methodology with particular reference to study design', *British Educational Research Journal*, 24(3): 255–68.

Goldstein, H. & Heath, A. (eds.) (2000), *Educational Standards*, Proceedings of the British Academy, vol. 102.

Goldstein, H., Huiqi, P., Rath, T. & Hill, N. (2000), *The Use of Value-Added Information in Judging School Performance* (London: Institute of Education).

Goldstein, H., Burgess, S. & McConnell, B. (2007), 'Modelling the effect of pupil mobility on school differences in educational achievement', *Journal of the Royal Statistical Society, Series A*, 170(4): 941–54.

Goldstein, H., Kounali, D. & Robinson, A. (2008), 'Modelling measurement errors and category misclassifications in multilevel models', *Statistical Modelling,* 8(3): 243–61.

Goldstein, H., Carpenter, J., Kenward, M. & Levin, K. (2009), 'Multilevel Models with multivariate mixed response types', *Statistical Modelling*, 9(3): 173–97.

Goldstein, H., Harron, K. & Wade, A. (2012), 'The analysis of record-linked data using multiple imputation with data value priors', *Statistics in Medicine*, 31(28): 3481–93.

Goldstein, H., Carpenter, J. R. & Browne, W. J. (2014), 'Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms', *Journal of the Royal Statistical Society, Series A*, 177(2): 553–64.

Goldstein, H. & Harron, K. (2015), 'Record linkage: a missing data problem', in Harron, K., Goldstein, H. & Dibben, C. (eds.), *Methodological Developments in Data Linkage* (Chichester: Wiley), chapter 6.

Gray, J. (1998), 'The Contribution of Educational Research to the Cause of School Improvement', Professorial Lecture, Institute of Education, London University.

Gray, J., Jesson, D. & Jones, B. (1986), 'The search for a fairer way of comparing schools' examination results', *Research Papers in Education*, 1(2): 91–122.

Gray, J., Jesson, D., Goldstein, H., Hedger, K. & Rasbash, J. (1995), 'A multi-level analysis of school improvement: changes in schools' performances over time', *School Effectiveness and School Improvement*, 6(2): 97–114.

Healy, M. & Goldstein, H. (1976), 'An approach to the scaling of categorised attributes', *Biometrika*, 63(2): 219–29.

Jacobs, P., Brunton, M., Court-Brown, W. M., Doll, R. & Goldstein, H. (1963), 'Change of chromosome count distribution with age evidence for a sex difference', *Nature*, 197(4872): 1080–1.

Joshi, H. (2020), 'A tribute to Harvey Goldstein', *Longitudinal and Life Course Studies*, 11(3): 285–7.

Leckie, G. & Goldstein, H. (2009), 'The limitations of using school league tables to inform school choice', *Journal of Royal Statistical Society, Series A*, 172(4), 835–51.

Leckie, G. & Goldstein, H. (2017), 'The evolution of school league tables in England 1992–2016: "con-
          textual value-added", "expected progress" and "progress 8"', *British Educational Research
          Journal*, 43(2): 193–212.
Levy, P. & Goldstein, H. (eds.) (1984), *Tests in Education: a Book of Critical Reviews* (London: Academic
          Press).
Longford, N. (1988), *VARCL—Software for Variance Component Analysis of Data with Hierarchically
          Nested Random Effects (Maximum Likelihood)* (Princeton, NJ: Educational Testing Service).
Mortimore, P. (2020), 'Harvey Goldstein Obituary', *Guardian*, 28 April.
Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988), *School Matters: The Junior Years*
          (Wells: Open Books).
Steele, F., Goldstein, H. & Browne, W. (2004), 'A general multilevel multistate competing risks model for
          event history data with an application to a study of contraceptive use dynamics', *Statistical
          Modelling*, 4(2): 145–59.

Note about the authors: John Gray is Emeritus Professor of Education at the University
of Cambridge, and former Vice-Principal of Homerton College; he was elected a
Fellow of the British Academy in 2000. Sir Ian Diamond is the UK's National
Statistician, and former Principal and Vice-Chancellor of Aberdeen University; he
was elected a Fellow of the Academy in 2005. Fiona Steele is Professor of Statistics
and Head of Department at the London School of Economics and Political Science;
she was elected a Fellow of the British Academy in 2009.