"After Fifteen Thousand Hours, where do we go from here?"

- Jeff Evans (Middlesex Polytechnic, Enfield) for the Radical Statistics Education Subgroup.*

After Fifteen Thousand Hours (FTH), where do we go from here? In response to this question, rather than dwelling on criticisms of the work by Ruther et al., we aim to see how· a scrutiny of it plus the various commentaries can stimulate ideas about what should now be researched, and how to research it. This scrutiny involves a critical awareness not only of the methodology, but also of the theoretical and practical context of their work, and of its use.

1. The history of quantitation research into the determinants of educational attainment, and in particular into school effectiveness, over the last 15 years or so, could be seen as a succession of ships passing in the darkness, each with its own fairly distinctive (conceptual) map, and each arriving at a somewhat different destination (in terms of theoretical conclusions and policy recommendations). Thus, in Table 1, we can see that the "main" independent variable - in the sense of the one found to be most significant statistically, or put forward as the most significant in policy terms - is different in each study. Furthermore, the variable put forward as the "main" one in Plowden, parental attitudes, was not even measured in most of the remaining studies.

Table 1 - Comparison of Quantitative Studies of Educational Attainment.

	"Main" Independent Variable.	Level of Schooling.	Unit of Analysis.	Sampled Population.	Researching Agency (Funding).	Type of Publication.
Plowden (1967)	parental attitude	primary	pupils/ schools	England	Offical (DES)	H.M.S.O.
NCDS (1972; etc.)	socio-economic class/parental involvement.	primary; secondary.	pupils	Great Britain.	semi-official (various Depts).	various.
Byrne & William- son & Fletcher (1975).	resources/ provision	secondary.	LEAs	England & Wales.	academic (S.S.R.C.)	"academic"
Bennett (1976; 1980).	teacher style	primary	pupils; class- rooms.	The North- West.	academic S.S.R.C.)	"popular"
Rutter et.al. (1979)	school process/ 'ethos'.	secondary	school	I.L.E.A.	academic (DES)	"popular"

Looking down the first column in Table 1, we are prompted to wonder whether each new study has brought successive increments in understanding - or simply a "new" independent variable. Now, one thing all these independent variables (except for socio-economic class) share is the capacity to be used as policy tools in the relatively short term; perhaps this constant change in the "main" independent variable is a reflection of disappointments in policy initiatives, or of developments elsewhere, esp. in North America (see Dave Reynolds' paper later in this symposium) or of changing intellectual or ideological trends. In any case, has such a succession of studies brought progress, either in policy, or in theoretical terms?

To be fair, the range of independent variables that can be studied in a given project depends on the resources provided for the project-plus whether a particular independent variable actually varies in that context (the "sampled population" in Table 1). Thus the cost of inclusion of a parental attitude or involvement variable in FTH may have been prohibitive. Also, if the levels of resource provision to each school were roughly the same in all ILEA secondary schools, there is little point in including that variable. (However, in a situation where heads can choose different mixes of different sorts of resources, e.g. trade off equipment purchases against school outings, it might be worthwhile studying the possible effects of the design of these choices). These two practical considerations are relevant to FTH; they do not mean that the variables parental involvement and level of resource provision (or the others highlighted in Table 1) should not be part of an overall theoretical framework developed and referred to by those researching school effectiveness.

Though certainly not wanting to recommend the establishment of one priviledged theoretical framework, we would suggest that whatever conceptual "map" is being used by a group of researchers should be made explicit, and acknowledged. Any research is shaped - and the data limited - by the way aspects of social life are selected as "problems", and the way they are conceptualised. Thus, whether because of conceptual preference or pratical constraint, neither level of resources provision nor parental involvement are on the conceptual map used in FTH; this

is fair enough, as long as it is acknowledged that this project itself can only be 'silent' on the issue of the effects of these variables. Yet Rutter et al. seem to want to claim that their work shows that "the main source of variations between schools in their effects on children did not lie in factors such as buildings or resources" (pp.179-80; our emphasis). This error - a most serious one in the current political climate - might have been avoided if their conceptual map had been more explicitly acknowledged.

Secondly, we can point to the usefulness of devoting effort to producing synopsis of programmes of already accomplished research, in order to ascertain the extent of "hidden continuities" for example, and re-analysing specific projects (see 10 below), as well as starting new ones.

2.   We can look at the conceptual maps of previous studies more systematically. In particular, once a new concept is "produced", conceptual and empirical work is needed to develop it. Thus with school 'ethos', we should ask questions such as the following about its conceptualisation and operationalisation:

-    Where did it come from, in terms of the researchers' previous conceptions?

-    To which earlier concepts in this research programme - or others - is it meant to relate - if any? For example, Sharp and Green characterise the 'school ethos' of the primary school they studied as "an identifiable ideology about its role and practices in relationship to its clients" (1975, p. 97).

-    Is it claimed to replace or subsume earlier concepts; e.g. Finlayson's 'school climate' (1973) or Plowden's school processes?

-    How many dimensions, or components, is it thought to have, and how are they considered to be related?

-    Could the construct validity of the indicators for each component of 'ethos' be improved?

-    How well do the different components of 'ethos' inter-correlate, and are they independent of indicators for variables claimed to be different (questions of convergent and discriminant validity)? For example, can indicators such as library use, outings, and clerical help be kept in an index for a variable like 'ethos' which is claimed to be independent of resources, - especially if attempts are to be made to apply the 'ethos' notion more widely (i.e. to schools representing a range of levels of resource provision)?

20

- Why are any particular weightings of the various components used in the overall index of 'ethos'?

- How much, if at all, do the conclusions of the study change when these weightings are changed?

There is still work to be done on these sorts of questions, in connection with 'ethos' Not only is exploring such questions good research practice, but also, having done so, the researchers (and the readers) are better able to respond critically to policy interpretations of the research's conclusions. (see 9 below).

3. Some aspects of the exploration of concepts need to be carried out through interaction with those individuals whose actions are being described and explained, i.e. using ethnographic methods. This is necessary for several reasons. First, there may be ambiguities and different perspectives on the meaning of certain categories; Andy Hargreaves mentions varying definitions of 'homework' (e.g. across subjects) and negotiations between pupil and teacher over 'lateness' (see his review of FTH).

Second, clear and coherent conceptual maps, "backed" by correlations which are statistically significant, can only give us "black box" insights into the relationships of interest – even if the correlations are produced wichin experimental designs! Thus, we need studies which describe empirically the processes by which school ethos (or whatever) is actually linked with various effects[1], through the consciousness of pupils (see Dave Reynolds' review).

4. The pinpointing of several dependent variables in FTH, and the comparing of their patterns of relationships with the independent variables, is to be welcomed. Of course, it alerts us to the possibility that the processes determining different dependent variables may be qualitatively different. For example, a radically different 'ethos' might be related to success in assessments with a significant project component, or to enhancing the relations between school and community. All this tends to suggest that developing powerful theories and broad research programmes is linked with the promotion of varied and innovative educational practices.

21

5. We return to the practical and methodological problems to do with lack of variation within a sample on some factor of interest. For example, if level of resource provision doesn't vary across schools in an ILEA sample, then variation could be located - or produced - in one of several ways:

(i) by using "naturally-occurring" variation in replications including LEA's and schools with different levels of provision;

(ii) by using variation over time, induced by cuts in provision, as a type of quasi-experiment; this sort of study is of paramount political importance;

(iii) by using a controlled experiment which randomly assigned different levels of resources - say, above some specified minimum - to different schools or LEA's. This sort of idea is certain to meet political opposition - and for good reason - in periods of financial constraint; it might be defensible during periods of expansion. (For ideas on situations where controlled experiments and quasi-experiments might be used, see Donald Campbell's "Reforms as Experiments" (1969) and other writings.)

In terms of studying the effect of variations in school ethos, one of us has wondered elsewhere if the idea of randomly allocating children to secondary schools is really as outrageous as it first seems (Ian Plewis in Barbara Tizard et al.)

Another way of experimenting (in the 18th Century sense of "putting into practice") without random assignment would be a followup study of schools where there was a decision by teachers to improve the "ethos" (as suggested by Jenny Hewison in Tizard et al.). Thus the so-called "scientific" question of whether school ethos has been demonstrated to have an effect on performance and other variables would give way to the more "political" question of whether the evidence is considered strong enough to justify such an intervention. There remains a more basic political issue: how will such a decision be made?

6. We turn now to the use made of educational research, beginning with its presentation and reception. The final column in Table 1 compares the type of publication for the five projects mentioned earlier; this and the style of presentation generally have an influence on how research is received by members of the public, and especially by interested parties: teachers, parents- and what about pupils?

In the case of teachers, the problems are substantial, often they are not aware of educational research findings; when they are, they are confused by them, or indifferent, or threatened (Ted Wragg, "Tactical Evasions on the Chalk-Front Line, Education Guardian, Sept. 1979). Many of the problems of course may have to do, not with communication skills used in research reports, nor with the quantitatitive competences of teachers, but rather with preceived political interests. Here this association's (BERA's) commitment to involving teachers much more extensively, both as "consumers" and as producers of research, is relevant. Put slightly differently, it suggests involving teachers much more fully in the formulation of research problems, and in the critical interpretation of results – rather than simply the reception of "findings". Steps in this direction might not only enhance the quality of research (see esp.3 above), but also broaden its audience.

7.   One feature of reports of quantitative research which makes them incomprehensible to most people is their use of "statistical metric_s", such as probability levels or proportions of variance "explained"– these are limited in the meaning they can convey compared to differences expressed in substantive terms, such as number of days absent, marks on a particular test[2], etc. We would endorse the calls (Jenny Hewison in Tizard et.al; Derrick, 1977) made to present results in terms of substantive differences between groups, in addition to – or instead of – the presentation of significance levels and amounts of r^2 . (For further discussion of the limitations of significance testing, see Carver, 1978 and Atkins and Jarrett, 1979).

8.   This strategy of simplifying statistical presentations may also require long term "competence-building" for its success. For example, analysis using ordinal or interval scale variables will normally be most simply presented via correlation coefficients, which are more difficult to understand than differences among groups. We would support initiatives from BERA to promote research methods and statistical competence amongst educational researchers – or among teachers. This might be done through writing mono-graphs, holding day schools, etc.

Competence-building can also be aided through the presentation of a statistical appendix to research reports. It is difficult to get the level right: is it for the average teacher, or for the relatively numerate one? This could be gauged by presenting a draft of such an appendix to a group at a Teachers' Centre or a B.Ed class, and responding to their comments. The level of the appendix on the log-linear model for FTH was probably too ambitious, an alternative, somewhat longer, attempt has recently been prepared by one of us (Dougal Hutchison, 1980).   23

9.   At many points in FTH, Rutter et al. put qualifications and caveats, e.g. about causality, on their interpretations of the research. However, there is widespread concern about the "hype" - oversimplification and dropping of qualifications - which has surrounded publication and discussion of their research. "Hype" can be produced at various stages: in the summary of findings in the last chapter, on the cover blurb, in summaries in the media, and in discussions of policy "implications" of the research.

Now of course policy recommendations are not unambiguously implied by the 'findings' (theoretical conclusions) of any research. Yet researchers should be sensitive to the sorts of policy recommendations which are likely to be pressed on the basis of their findings, and they could comment critically on whether - and under what conditions - the research would in their judgement justify such recommendations. This would have the extremely salutary effect of making the researchers put their qualifications and caveats "into practice", and should enable them to help moderate the "hype" in cases where it is likely.

In the case of FTH, we could ask the researchers to comment critically on the extent to which their research justifies each of the following policy recommendations:

(a)   In schools where, for instance, exam performance is below average, the head should personally take charge of improving the "school ethos" by tightening up the authority structure (as might be implied by the term 'tight ship' used in the TES, 13 June), - in order to check better on punctuality, consistency of discipline and homework.

(b)   Schools with a particularly good ethos should not be amalgamated, even with smaller schools, in the current restructuring, since that would mean the loss of their ethos and style, qualities wanted by parents. (The Education Secretary's justification for not amalgamating Highbury Grove School with another; reported by John Fairhall, Education Guardian, 13 May, 1980).

(c)   Parents seeking to choose a comprehensive school for their child should ascertain the school 'ethos' by paying it a visit (Maureen O'Connor, Education Guardian, 5 August).

10.   Nor do data and the results of statistical analyses unambiguously "imply" the theoretical conclusions or findings drawn from them. Thus other researchers need to be able to re-analyse the data using their own theoretical-frameworks – as far as this is possible – or different thchniques of analysis.[3]

An example of a possible re-analysis using the F TH data: some of the indicators for school ethos might also be acceptable indicators for the allocation of resources among certain crucial uses within the school; for example, library use, outings, pupil conditions, decorations of classroom, clerical help, perhaps teachers' time spent on assembling equipment, and so on (see Appendix E).

Thus, the data from F TH might be useful for studying the relationship of the mix of resources to pupil attainment, and to other dependent variables of interest.

Therefore, we look forward to the data from Fifteen Thousand Hours, and other major studies of educational attainment, being deposited in the S.S.R.C. Survey Archive, as a matter of course.

We can summarise our recommendations as follows (the numbers refer back to sections in the main text):

1.   the making explicit of the 'conceptual map' used in a particular study, in order to avoid pitfalls of 'crude empiricism";

1a.   synoptic work on programmes of already-accomplished research, in order to ascertain the extent of "hidden continuities" and re-analysis of particular projects;

2.   developmental work on new concepts such as school ethos;

3.   study of the processes relating variables, through research with an ethnographic component;

4.   attention to overlooked areas, such as performances on innovatory modes of assessment;

5.   Consideration of experimental and quasi-experimental designs, as resources for locating or producing variation on certain "difficult" factors;

6. consultation with teachers concerning the formulation of research problems and the interpretation of results, with a view to enhancing the quality of research and broadening its "audience";

7. Use of "substantive metric s" in addition to – or instead of – "statistical metric s" in the presentation of results;

8. BERA initiatives towards "competence building" amongst educational researchers, and among teachers.

9. minimisation of "hype" through researchers' "control" of possible policy recommendations;

10. reanalysis of previous projects, using different theoretical frameworks (or different tools of analysis)

10a. facilitation of re-analysis, through the depositing of data from major studies with the S.S.R.C. Survey Archive, as a matter of course.

NOTES.

1. An example of a large scale study which incorporated an ethnographic component – the 'CAI' evaluation – is described in Evans, 1979.

2. Some would query even the relative meaningfulness of many of the substantive measurements used in psychology and education; e.g. "seldom have we collected data about the measurement scale that would allow us to determine whether the new-score difference of ten points is scientifically significant or not." (Carver, 1978, p.393; our emphasis.)

3. As an example of the latter, see Cherkaoui's re-analysis of the first IEA data, using two different models.

References.

Atkins L. and D. Jarrett (1979), "The Significance of Significance Tests", Ch.8 in in Irvine J., I. Miles and J.Evans eds. Demystifying Social Statistics, Pluto Press.

Campbell D.T. (1969), "Reforms as Experiments", American Psychologist, vol.24, pp.409-29.

Carver R.P. (1978), "The Case against Statistical Significance Testing",
Harvard Education Review, vol.48, no.3, pp.378-99.

Cherkaoui, M. (1979), Les Paradoxes de la Réussite Scolaire, Pressed Universitaires de France.
(currently being translated into English, for RRP).

Derrick T. (1977), "The Criticism of Inferential Statistics", Education Research, vol.19, no.3,
pp.35-40; see also rejoinders in vol.19 no.4 and vol.20 no.1, esp. that by
G. Chanan.

Evans J. (1979), "Evaluation of Research Design", Block 3, Part 6 of DE304 : Research
Methods in Education and the Social Sciences, Open University.

Finlayson, D.S.(1973), "Measuring 'School Climate'", Trends in Education, April 1973.
pp.19-27.

Hutchison D. (1980), "Statistical Appendix", pp.235-265 in J. Steedman, Progress in
Secondary Schools : Findings from the NCDS, National Children's Bureau.

Pawson, R. (1978), "Empiricist Explanatory Strategies: The Case of Casual Modelling",
Socidogical Review, August, pp.613-45.

Radical Statistics Education Subgroup (1980), "What constitutes responsible statistical
criticism?" and "Sketch for a Demystifying Approach to Social Research",
Radical Statistics Newsletter, 18, pp.15-31, (available from: Radical
Statistics Group, 9 Poland St., London W.1.).

Reynolds D.,A. Hargreaves and T. Blackstone (1980) "Review Symposium : Fifteen Thousand
Hours", Br. J. of Sociology of Education, vol.1, no.2, pp.207-219.

Sharp R. and A. Green (1975), Education and Social Control, RKP.

Tizard, B. et.al. (1980) Fifteen Thousand Hours : a Discussion", Univ. of London Institute of Ed.

Zuniga R.B. (1975), "The Experimenting Society and Radical Social Reform", American
Psychologist, vol.30, no.2, pp. 99-115.