

15,000 Hours - Summary by Ludi Simpson

15,000 Hours - published by Open Books, 1979. Financed by ILEA, directed from the Institute of Psychiatry, University of London. Design: 12 Inner London Education Authority secondary schools, selected to be representative of different school types. Classroom observation, teacher interviews, and school records used in data analysis.

INDEPENDENT VARIABLES

Intake characteristics

(Social class; verbal reasoning group at age 11)

Physical characteristics

(Size, age, type, site of school)

46 "Process Variables" (eg . emphasis on homework, sports prizes available, school uniform, degree of discipline). 39 of these were also summed to give one overall "process" variable, labelled the schools "ethos".

DEPENDENT VARIABLES

- a) Attendance
- b) Attainment
- c) Behaviour
- d) Delinquency.

MAIN CONCLUSION - "Not only were pupils influenced by the way they were dealt with as individuals, but also there was a group influence resulting from the ethos of the school as a social institution..... Schools can do much to foster good behaviour and attainments, and even in a disadvantaged area, schools can be a force for the good."

Criticisms

1. Non-random sampling. 12 schools were selected to be representative of different school types. It is difficult to see to what population (if any) these can be generalised, so the use of statistical significance tests, which aim to determine whether a result in a sample can be generalised to a population, is highly suspect. I feel that it would be legitimate to generalise the results to situations which resembled the study one. This would presumably mean other ILCA schools, but even other inner city schools or even other inner city schools in the UK. Is this not too harsh? Surely, there is no 'real danger of invalidating the results - SR2? Well, common sense or a 'gut feeling' may make one wish to generalise further, but remember that ILCA is far better-off than other UK inner-city authorities, so the conclusion that the book reaches that fluctuations in resources in a well-supplied authority make little difference cannot be generalised to argue that expenditure cuts elsewhere do no harm.
2. Significance tests were in addition used in an extremely Procrustean way - results with $P .05$ being considered significant and reported and those with $P .05$ being ignored. With a sample of 12 (or upon occasion, 8) this leads to some really rather large differences being ignored. See P.99 where a difference in average ranks on a sample of 12 of 4.0 to 7.8 is described as non-significant though it is about as large as could arise on a sample of that size. Also significance levels are quoted rather than estimated fitted constants.(SR11).
3. The 'ethos' variable is dubiously constructed and dubiously interpreted
The items which were used to construct it were selected out of a larger set solely on the basis of the significance of their correlation with the outcome, rather than on substantive considerations such as the relevance to the measurement of some aspect of a sociological or educational theory. How serious is this, given that 39 of the original total of 46 indicators were actually used and only 7 discarded. Commonsense suggests that such a procedure might not seriously distort results but rather 'highlight' them so that they appeared slightly more statistically significant than they otherwise might, but Pries (1949) has shown that such a procedure can give rise to apparently statistically significant results from data from tables of random numbers. My own feeling on this is that Pries's finding is rather surprising - could it be the 1 in 20 chance of a

false positive on a test? In any event this underlines the necessity for studies of robustness of statistical techniques. Jeff Evans in the next part of this discussion also criticises the interpretation of the collection of variables which go to make up the 'ethos' variable, as well as the conceptualisation of the variable.

4. There was to be little if any check on the validity of the measures used.

Surely it not enough simply to ask a teacher how he/she would deal with misbehaviour by '10 Naughty Children'. In such a 'sensitive' area some kind of comparison with actual behaviour, or behaviour as reported by (for example) pupils would be necessary for at least a sample. This type of problem of course occurs in almost any questionnaire-based research, and has been the subject of much investigation, but Rutter et al seem to ignore this. Answers could tend to be more representative of what the teachers would like to do, rather than what they do to, their ideals rather than their behaviour.

5. The unit of analysis (the school) hides within school differences relating to a school's overall' performance. Likewise the use of the school mean hides the fact that a very important outcome of schooling is a 'spread' of attainment within the school considerably greater than between school mean differences.

This is of course pretty well standard in such research, and it would not be reasonable to pick on 15,000 Hours specifically for this.

6. While the authors recognised the importance of controlling for each school's intake, this control as carried out in the project was inadequate

Ignoring the question of whether the right intake variables were measured, the researchers seem to have unnecessarily weakened the power of their firstly by using them one at a time, and secondly by changing the coding of the pretest score from 1 categories to 3, thus both weakening the predictor and giving uncomparable results. (Perhaps the 3-category coding came from a different source, but a discussion of this is necessary). If control for pre-existing intake differences is inadequate, as I believe it was, than artefactual school progress differences can arise. This renders the very existence of interschool differences suspect which in turn of course undermines the lengthy explanation of the causes of these differences.

7. 'Alternative or complementary explanations. As well as school ethos or organisation

- (a) Personalities and experience of teachers, in particular the head teachers are of particular importance
- (b) The ecological effect of the other pupils in the school in creating a particular atmosphere

These (a) and (b) will contribute with school organisation/ethos towards success or otherwise. One is always dubious about putting forward such alternatives but any result of these would contribute a 'school effect' as defined in the book and I suspect that parents at least would consider them to be more important. Yet they are not measured or mentioned by the researchers.

There is nothing purely statistical about these criticisms - perhaps the only justification for a statistician to put them forward would be if only statisticians were capable of penetrating through the numberage sufficiently to notice the gaps.

8. The statistical methodology was not well explained and so difficult to check. In particular the statistical appendix would be incomprehensible to anyone except a statistician, and thus be of no benefit to the teachers, administrators and parents at whom the book was presumably aimed. On the other hand it would tell a statistician nothing new.

I feel strongly that the ideal role for a Statistical Appendix is to give and explain the technique used to those reading the rest of the book, if at all possible. With a standard technique, such as the log-linear analysis used in the book a heuristic and intuitive description should suffice, together with a reference to a standard article or (preferably) textbook. Any new techniques or theories derived should also be described, and here one may have to be more mathematical.

In 15,000 Hours however the technique and its applications were quite standard and the appendix could serve only to mystify readers.

9. Twelve case studies would have been equally useful. The design was poor, falling uncomfortably between case-study and survey. 12 units is really too small a number for statistical analysis and more must have been lost in the attempt than could be gained by the use of statistical techniques. It is difficult to escape the conclusion that the only benefit of the statistical procedures was to give an air of respectability.
10. Secondary analysis. To gain access to the schools, the research team had to obtain the permission of ILEA. The latter have refused on grounds of confidentiality to allow secondary analysis of the data. If research cannot be checked, it's useless in helping ascertaining the truth; This sensitivity of administrators and teachers in the long run will seriously discredit research - not only will it be useless but it will be seen to be useless.

A similar problem arose with N. Bennett's Teaching Styles and Pupil Progress: the researcher originally declined to allow other researchers to have access despite the fact that one of the teachers had appeared on television with her class!

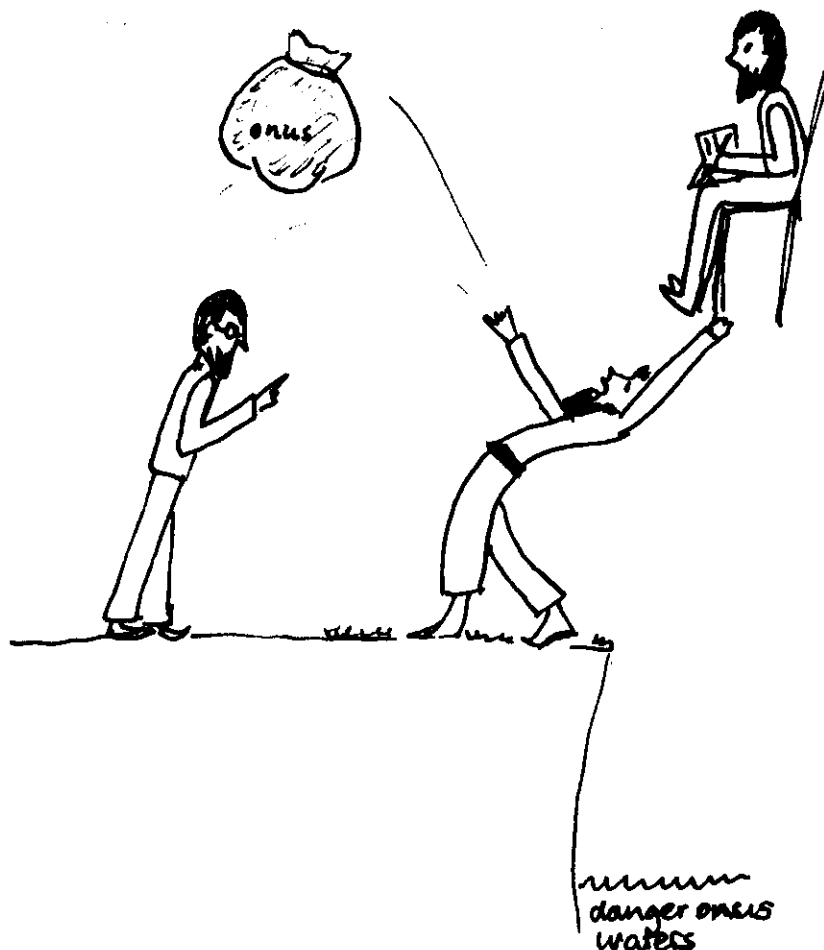
11. Critics have commented on what might be succinctly described as 'hype', i.e. promotion of a product beyond its intrinsic merits. To quote Wedge (1979).

'The second count concerns ambiguities in the presentation. Rightly, the project team report tentatively a "strong inference" that "school life had truly influenced children's attendance, behaviour and attainment" and stress that this was "not a direct demonstration of an influence". But one also reads seemingly more bold assertions: "Much of the effects of secondary schools were linked with their features as social organisations". Or again, "our findings suggested that pupils behaved better and achieved more when teachers treated them in ways which emphasised their successes and potential rather than those which focussed on their failings and shortcomings". To the sophisticated reader of research reports, this last statement contains two clues that a neat cause-effect relationship has not been established. But those "professionally involved in teaching" or "with a care for the quality of education today" for who the publisher tells us this report is 'vital', are not in general sophisticated readers of research reports. Thus, particularly for the practitioner, regrettably, the report disappoints because it is not always sufficiently explicit and hence is confusing. Paradoxically, at times the techniques used are too complex to be simply expressed for lay reading, but the authors include them.

Wedge's comments seem to me to have substance. However, they are comments which do not need a statistician to make them. Wedge is not a statistician.

References

- BENNET S.N. (1976) Teaching Styles and Pupil Progress Open Books.
- BROSS, I (1960) 'Statistical Criticism', J A S A. 394.
- PRATT, T. BURGESS, J. ALLEMANO, R. & LOCKE, M. (1973) Your Local Education Penguin.
- PRIESS (1979) '1500 Taus and Rhos' Research in Education.
- RUTTER, M., MAUGHAN, B., MORTIMORE, R., OUSTON, T. (1979) Fifteen Thousand Hours. Penguin
- WEDGE, P. (1979) '15000 Hours: Review', Disadvantage in Education, Summer 1979.



Bross puts the onus on the critic while leaning over backwards to support the writer.....