

Teacher Effectiveness and the Explanation of Social Disparities in Educational Achievement

New Zealand Journal of Teachers' Work, Volume 1, Issue 1, 42-51, 2004

ROY NASH

Massey University

ABSTRACT: The belief that effective teaching can raise the performance of students to a marked extent seems to have become fixed in the contemporary discourse of educational policymakers. It is asserted that evidence-based research has demonstrated a causal relationship between better teaching and better learning. A critical examination will show, however, that the case is less secure than it might seem, even within the logic of quantitative modelling. The argument is based on an examination of Hattie's influential meta-analyses of teacher effectiveness research and on an analysis of some relevant data on teachers' expectations from the New Zealand PIRLS 2000 dataset.

Research indicates that effective classroom teaching can explain up to half of a child's educational achievements. (Mallard, 2003)

What does it mean to assert that, "effective classroom teaching can explain up to half of a child's educational achievements"? I put this question to a 200-level sociology of education class. The students were puzzled. After some discussion they reached the hesitant consensus that about half of what a child knows might have been learned from effective teachers. It seemed unkind to then ask them whether the other half might have been learned from ineffective teachers! The very fact that the proposition "half of what a child knows is learned from effective teaching" seems to make substantive sense to students of education should be a matter for concern. What programme of research would be needed to establish such a finding? Would it not be necessary to list all that students knew and to allocate each item, by some more or less reasonable procedure, to the categories 'learned from effective teaching' and 'not learned from effective teaching'? The whole idea is absurd. Yet, perhaps one should not be surprised by the students' interpretation. The Minister's statement is actually incomprehensible to anyone without some statistical knowledge, but its construction manages to leave non-critical readers with the general impression that effective teaching makes a substantial contribution to a child's learning, and that is likely to be the point in the context of its release. In order to make sense, however, the key words should read "effective teaching can explain up to half the variance accounted for in educational attainment". But as that statement is expressed in a technical jargon quite unsuited to the style of a press release it has no doubt been redrafted into something more like plain language. The result is a sentence with a meaningful semantic structure – one with the sense that good teachers are responsible for half of what a child

achieves at school – that manages to convey a profoundly misleading conception.

The effort to explain what the Minister's statement really means will require, first, a discussion of how statistical modelling attempts to allocate weights to the several 'factors' involved in the production of differences in educational achievement; second, an examination of Hattie's recent contribution to this area; and, third, to a critical comment on research into teacher expectations drawing on empirical data from the recent New Zealand PIRLS 2000 data (Gonzalez & Kennedy, 2003). It is my aim to show, within the constraints of a non-technical and practice-focused paper, that educational research should never be excused critical attention. The process of unpacking the Minister's statement may take a little time.

THE EXPLANATION OF VARIANCE IN EDUCATIONAL RESEARCH

According to a recent British study, the extent of the difference between the level of cognitive development reached by the most and least successful 12-year-olds is about 10 years. In the author's own words: "The range of mental development is far, far wider than anyone dreamed ... in a representative sample of 12-year-olds around 7 per cent test at the level of the average 6-year-old, and 10 per cent at the level of the top 30 per cent of 16-year-olds" (Larkin, 2002, p.191). This assessment is based on Piagetian tests and there is good evidence that the range is almost equally great when measured by standardised tests of educational achievement. A precise quantitative expression can be given to the amount of variance in a set of achievement scores. But, what does it mean to 'explain' the variance? When statistical analysts speak of 'explaining' or 'accounting for' the variance, or of 'allocating x percent of the variance to factor y', they mean that an indicator has been found that correlates with the scores they want to explain, and the magnitude of the association enables an estimate of the variance explained to be calculated. The correlation coefficient extends from -1 (perfect negative correlation), through 0 (no correlation) to + 1 (perfect positive correlation).

In educational research most correlations tend to be moderate to low and a few representative examples may help non-statisticians to form a sensible impression of this statistic: IQ test-retest, about 0.9; between mathematics and science tests, given at much the same time, perhaps 0.7 or 0.8; between social class and achievement, about 0.3; and between aspiration and achievement in primary school, about 0.2. It is a simple matter to work out a correlation coefficient (no one does them by hand any more), and the variance explained is just the square of the correlation. When analysts say that SES (socio-economic status) explains about 9% of the variance in educational achievement, they reach this conclusion because the correlation between SES and achievement is usually about 0.3. The statistical calculations involved are elementary – they can be found in any introductory statistics text – and it requires only a little specialist training to grasp the logic of the mathematical arguments, but the status of the explanations provided by statistical modelling raises very much more complex problems.

The assertion that effective teaching can explain up to half the variance in educational achievement should now become a little more clear. To say that half the variance is explained is to say that the range of correlations reported between indicators of effective teaching and educational achievement extends to 0.7. This should be context enough to show how misleading is the statement that “half of what a child knows is explained by effective teaching”. Statistical models have nothing whatever to say about individuals, and to know that an indicator accounts for half the variance in the achievement of a group provides no information about the causes of individual learning. It is a bizarre fallacy to suppose that the statement “up to half the variance is explained by effective teaching” entails the conclusion that “up to half of a child’s educational achievements can be explained by effective classroom teaching”. This does not mean, of course, that the statement when correctly interpreted should be accepted on its own terms. This may be an appropriate point, incidentally, to note that the phrase “up to”, while possibly correct, may have the potential to mislead. It is a bit like saying that newborn infants can weigh up to 6 kilograms: they can be that heavy, but the median weight is only about half that. The range of values for the variance explained by teacher effectiveness, or for any characteristic of teachers, may be between 0% and 50%, but the median value is well towards the lower end of the range. This is a point worth noting – but there are more important matters to consider.

The most fundamental problem for quantitative modelling is that of explanation. When the variance in a set of scores is said to be “explained”, or “accounted for”, by a correlation this is, of course, a technical usage restricted to the statistical model. It is all a matter of internal definition. The crucial thing is the nature of the relationship between a model and the reality it supposedly represents. This is something that can scarcely be considered within the given framework of statistical modelling. It is deemed sufficient that ‘valid’ and ‘reliable’ indicators have been obtained and that some more or less *ad hoc* theoretical account of the relationship between ‘variables’ can be provided. In the conventions of the paradigm these minimal conditions allow it to be supposed with little more ado that no distinction need be made between the statistical model and the social mechanisms that generate the data.

Correlations are not causes. All statistical texts explain, usually in the first few pages, that the presence of a correlation does not necessarily imply the existence of a causal relationship between the associated properties or events. The PIRLS data show a positive correlation, for example, between low levels of student reading achievement and teachers’ attendance at reading seminars, but we would be reluctant to conclude that teachers who seek to further their knowledge of reading in this way were less effective than others (Gonzalez & Kennedy, 2003). The correlation is most likely to reflect the fact that teachers with many poor readers in their classes are more likely than others to consider themselves in need of further study. And yet, if ineffective teachers were required to attend reading seminars then the correlation might well bear a causal interpretation in that direction. A correlation contains no information about its origins. It is evident, therefore, that attributions of causality made on the basis of statistical correlations are entirely dependent on information not formally included in the model.

Variability in educational achievement is generated by the actions of parents, teachers, and students in homes, schools, and peer groups whose actions flow in an endless braided stream of practices that must be checked, separated into distinct processes, and expressed in quantified indices. Statistical modelling aims to quantify the separate contributions to variance in educational achievement by providing an ordered and weighted list of the factors held to be important. One of the most influential exponents of this form of analysis in New Zealand is Professor John Hattie of Auckland University. Hattie's research seems to be held in high esteem in the educational community, and not least by the Ministry of Education. It will be useful to examine Hattie's position on teacher effects.

One of Hattie's most influential, and perhaps most easily accessible expressions of his views – the article was presented at a conference of the Australian Council for Educational Research and posted on the New Zealand Ministry of Education website – will reward our attention. Although Hattie (2003, p.3) invites us to take his “synthesis of over 500,000 studies” seriously, a sceptical reader may be allowed to raise an eyebrow. At the rate of 50 every working day it would take 45 years to read that number of studies and Hattie's analysis, which is plainly not the result of an actual meta-analysis, is probably best regarded as an informed estimate. The analysis identifies six variables, student, home, teacher, social class, school and principal. The interaction effects between these variables, according to Hattie, are usually minor and are ignored in his summary model. The largest proportion of the variance, about half, is accounted for by student characteristics (these include prior achievement, self-concept, aspiration, and so on), and this leaves the remaining half of the variance to be shared between the other variables.

The model, it should be noted, is already highly problematic. It is a theoretical decision to privilege prior achievement, which almost invariably accounts for most of the variance in achievement, before taking into account the effects of family environment where the cognitive skills largely responsible for school achievement may have been developed. Non-specialists in this area cannot be expected to understand that models of analyses of variance are highly sensitive to the order in which variables are entered into the equation. It matters a great deal to the results whether the model puts student variables before home environment or whether the order of those variables is reversed. The quantitative estimates of the proportion of the variance explained by each factor or variable included in the model are quite different, and the decision about the order of entry is theoretical rather than technical. The question is: ‘What model represents reality most accurately?’, and that requires close attention to theory. Hattie's model, having taken out most of the variance, then allocates the second largest proportion, about a third, to teachers. This amount is therefore higher than that of the measured influence of social class, peer groups, principals and schools all put together. Hattie (2003) can now reach his key conclusion: teachers, he asserts, account for about 30% of the variance: “it is what teachers know, do, and care about which is very powerful in this learning equation” (p.3).

A further point remains to be made. Hattie lists about 30 different factors, subsumed within the big six, that contribute to an improvement in test scores

and provides an estimate of their individual effect size. Hattie presents his findings in terms of effect sizes expressed in standard deviation units. If we want to know how much students have learned as a result of a new programme then we compare the mean scores of students who have experienced it with those who have not and, assuming that the students are alike in all relevant respects, observe that when assessed the mean scores of the former are four marks higher than those of the latter. Is a difference of four marks worth taking seriously? That depends on the size of the sample; in general, the larger the sample the higher the statistical significance is likely to be, but it is always useful to express the difference as an effect size expressed as a proportion of the standard deviation. This paper cannot be an instant guide to statistical ideas and it must suffice to describe the standard deviation as an indicator of the range of scores. If, say, 16 marks covers the middle two-thirds of the scores then the standard deviation is about 8, and four marks is thus 0.5 standard deviations. Hattie reports that the positive factors affecting student performance range from teacher feedback (effect size 1.3 standard deviations) to team teaching (0.06). Hattie does not suggest that the influence of these factors is independent and additive but the presentation may mislead non-technical readers. All of the factors mentioned will be inter-correlated, some of them quite strongly no doubt, and the idea that maximizing all the factors on the list would increase achievement levels by about 15 standard deviations is quite absurd.

Perhaps more importantly, this line of research requires us to accept, for example, that some teaching activity recognisable under an objective description of 'teacher feedback' can be identified, quantified, and isolated as a causal agent in the generation of learning. But that behaviourist assumption cannot be accepted without challenge. What teacher behaviour is to count as 'feedback', how such 'objective' behaviour is subject to contextual definition, and whether it should be the students' rather than the observer's interpretation that matters, are all questions of central importance. What might count as 'feedback' in the catalogue of behaviours prepared in the context of suburban Illinois will not necessarily have the same status in the multi-cultural schools of Auckland. The effects on students, moreover, of a behaviour described in the typology as 'feedback: type 1', or whatever, cannot be presumed to have the same causal effect on students in different educational systems. The very use of the term 'causal effect' in this context should indicate the nature of the problem. Students are not like billiard balls to be shot into this and that pocket as if the teacher were some kind of self-propelled cue ball. The classroom is a place where actions are negotiated, rather than behaviours displayed, and where the consequences of those actions, both short-term and long-term, are mediated by complex processes of cultural interpretation.

This critique should be taken seriously. It is not that statistical techniques, the very foundation of evidence-based research in the scientific tradition, should be abandoned. If there is good evidence that, for example, the incidence of heart attacks can be reduced by programmes directed at the prevention of obesity by diet and exercise regimes, then it would be foolish to dismiss it on the grounds just advanced. But the situations discussed are not remotely similar. In some respects we are like billiard balls; if we eat too much and exercise too little then we will get fat, but if we understand what might be

recognised as 'positive feedback' in suburban Illinois as 'a stink showing up' in south Auckland, then our responses, a supposedly invariant causal response to a supposedly objective stimulus, will actually not be the same at all. As there is no real defence to this criticism, many quantitative researchers simply ignore it, but that is no reason why everyone else should do the same.

Nye, Konstantopoulos and Hedges (2004), in a contemporary review of teacher effects research, present conclusions very different from those reached by Hattie. The median estimate of 17 analyses derived from 6 high-quality United States studies carried out in the period 1971-2002 is 0.11 with the range 0.07 – 0.21. The Minister's preferred value, it will be recalled, is 0.7. One of the most recent of these studies, based on a representative national sample and using the public NAEP data (National Assessment of Educational Progress), reports a value of 0.12 in mathematics (Goldhaber & Brewer, 1997). Nye, Konstantopoulos and Hedges' own analysis uses data from the STAR Project (Tennessee) and allows them to estimate that the most effective 25% of teachers might raise student achievement a half standard deviation above that of the least effective 25% of teachers. This appears to be a substantial magnitude. However, these researchers have very little idea of what teaching practices are actually responsible for this statistically detected effect and the possibility that these findings are, to some extent, an artefact of non-random teacher-student assignment cannot be discounted. It is one thing to identify teachers as effective on the basis of observations of their teaching and then determine the extent of their teaching on student achievement, and quite another to find that some students do better than others and, by a process of statistical elimination (always a poor proxy for genuine experimental control) to reach the default conclusion that observed differences in student achievement must be attributed to teachers. The STAR research adopted the latter rather than the former method. There is no reason to doubt that teachers' efforts must count for something, but it is unlikely that policies designed to raise teacher effectiveness will eradicate, or even much reduce, social disparities in educational achievement.

ARE TEACHERS' EXPECTATIONS LOW?

The theory that variation in educational performance can be eliminated by raising the expectations of teachers is widely held. Indeed, the theory that processes triggered by teacher expectations are a key mechanism in the production of under-achievement is taught to education students as if it were an established fact. There are actually many reasons to question this thesis. When children enter school at the age of five there are already sharp differences in literacy and numeracy skills associated with social class and ethnic origin. These differences obviously cannot be due to experiences at school, and it is unreasonable to suppose that they are caused by differential pre-school attendance or systematic variation in the quality of their pre-school education. But, if relative cognitive differences between groups are simply maintained as they progress through the education system then there seems to be no reason to accept that these are created by the beliefs and actions of teachers. And if group differences in achievement are not created by teachers, then why should

it be supposed that they can be eradicated by raising teachers' expectations? Is there any evidence, in fact, that teachers' expectations for working-class and ethnic minority students *are* low? What evidence there is suggests, on the contrary, that New Zealand teachers' expectations are rather high. The most recent data in this area is provided by the PIRLS research.

The PIRLS (Progress in International Reading Literacy Study) was carried out in 35 countries, including New Zealand, by the IEA (International Association for the Evaluation of Educational Achievement) in 2001 (IEA, 2003; Gonzalez & Kennedy, 2003). It provides information on the reading achievements of a nationally representative sample of students aged about 10 years. The study is particularly useful because the data, collected from students, teachers, principals, and parents by questionnaires, are freely available for secondary analysis. The New Zealand dataset includes 2,504 students in 172 schools. Students were tested in complete classes, but there are often very few students with the same class identification, particularly in low decile schools, and this may largely be due to the exclusion of students with little understanding of English. There are only 6 classes in decile 1-3 schools, compared with 19 in decile 8-10, with 20 or more students. The relevance of this point will become clear in the discussion. Achievement in reading was assessed using instruments in separate blocks designed to cover several domains of reading literacy, and a number of values are provided for analysis.

Theories that suppose that students attending low decile schools are affected by teachers' low expectations in such a way that they become disaffected and reluctant to learn, are not supported by the PIRLS evidence. Most students of this age appear to be content with their experience at school. The responses to items designed to tap students' perceptions of the school environment and teachers' attitudes towards them are given overwhelming endorsement. The percentages of European (and Maori) students who answer 'a lot' (the highest on a four point scale) to some key items will support this statement:

I feel safe when I am at school, 60% (67%)

I like being in school, 42% (60%)

I think that teachers in my school care about me, 61% (66%) and

I think that teachers in my school want students to work hard, 91% (87%).

It is worth noting, perhaps, that students in low decile 1-3 schools are somewhat more likely than others to indicate that they like school, 52% (64%), compared with 40% (60%) in decile 8-10 schools, and that teachers care about them, 67% (68%), compared with 62% (62%). These findings suggest that primary school students, both European and Maori, are generally happy at school and, if anything, rather more so at low decile schools than at high decile schools. It would be difficult to interpret these results as evidence in support of the hypothesis that teachers in low decile schools hold poor expectations for their students whether they are Maori or non-Maori.

Teachers' own expectations also seem generally high. When asked how many students in the target class they expected to grow up to be good readers, more than half checked the response 'all or almost all' and fewer than 10%

indicated that it might be 'about half' or 'less than half'. Teachers in low decile schools may be somewhat more likely to hold expectations other than the highest: in low decile schools 21% of students are taught by teachers with lower expectations than those in high decile schools (3%), and there is a tendency for more Maori (16%) than European students (8%) to be in classes where teachers express lower expectations. Nevertheless, it should be pointed out that half of all students in low decile schools, and 51% of all Maori students, are taught in classes where teachers believe that 'all or almost all' students can become good readers, and it seems implausible to suppose on the basis of the evidence that this characteristic of teachers can have a large independent effect on student achievement. There are so few PIRLS students in some 'classes' that one could not, in any case, assume that those a teacher with 'low expectations' believes are unlikely to become good readers were even included in the PIRLS assessment. In some classes it must be supposed that most students have actually been excluded, and in some of those cases, particularly as 'mainstreaming' is normal practice, that may actually be by reason of their intellectual incapacity. There is something a little unfair in labelling a teacher as one with 'low expectations' because she is apparently disinclined to assert that 'all or almost all' of her pupils can become good readers when some of them may have a diagnosed learning disability. Special needs students in this category were excluded from PIRLS, but not explicitly from the reference to the teachers' 'target class'.

This well-conducted international study should not lightly be dismissed. It is possible, of course, to express reservations. Teachers may be reluctant to report that they expect only 'about half', or even fewer, students in their classes to become 'good readers', for teachers understand well that this is a datum that can be used against them, and students may have their own reasons for responding cautiously to items that invite them to express critical opinions about their teachers' actions. As PIRLS has no longitudinal data the hypothesis that students in the classes of teachers with higher expectations make more progress than others cannot be tested. These caveats must be admitted, and yet the evidence does not support the expectations hypothesis and the implications of that should be faced, particularly in light of the fact that there is no reason to believe that teacher expectations either create social disparities in education or are the principal mechanism responsible for maintaining them.

When this research is set in context with Hattie's conclusion that the teacher is the largest non-student source of variance in achievement the wider implications become clear.

CONCLUSION

There is something quite dangerous about the use of quantitative research for propaganda purposes. It is likely that not one sociologist of education in ten is competent to critique statistical methods in their own terms, and it is unlikely that the proportion of teachers so equipped is any greater. Whether the Minister even understands his own announcement must be considered an open question. It certainly does not mean what it seems to mean, and when correctly understood plainly misleads by omitting to note that the upper range of teacher

effects is four or five times higher than the median value reported by leading studies. Perhaps it is unrealistic to expect a Minister to declare in such a context that “anything between zero and half”, and so on and so forth, but “up to half” is really not the most objective way to present these complex findings. It is hardly surprising that so many involved in education, including parents, teachers, students, policy makers, advocates, are so deeply suspicious of quantitative methods in educational research. Although objections to quantitative modelling can be overstated – for it would surely be an error to abandon the attempt to estimate the relative importance of separate agents and social organizations in the generation of variance in educational achievement – realist criticism of the positivist assumptions of explanations derived from statistical analysis are well-founded and should certainly prevent us from accepting them without close examination.

The official support, so evident in contemporary policy discourse, for the view that social disparities in education are to a substantial extent a product of teacher expectations is perhaps not unmotivated. Is this theory really likely to create a climate in which teachers are encouraged to develop and exercise their professional skills free of a level of intrusive supervision and monitoring? There may be a real chance that it might lead to a regime of a very different kind. If, as a recent *Sunday Star Times* editorial asserts, “research has found that improved teacher attitudes to pupils can result in the so-called dunces being turned into effective, learning students in a surprisingly short time”, then if the attainments of schools in Mangere and Porirua do not match those in Remuera and Fendalton, the teachers in the former areas are revealed as inadequate. If schools really are responsible for up to 80% of the variance in attainment, and if this is read as establishing the limits of schooling as an agent of equality, then the conclusion cannot be resisted. Teachers should not be required to accept as much as this. To impose so excessive a demand upon teachers is simply to exploit their professional goodwill and collective social conscience. Those who accept this position, and take upon themselves this burden, are acting with a larger generosity of professional spirit than may strictly be necessary. At least, they may not need to be over-impressed by statements that appear to be based on the ‘evidence-based’ research that so strongly influences official policy on the importance of teachers’ expectations. A press release that tells us that, “effective classroom teaching can explain up to half of a child’s educational achievements”, is in itself meaningless and derived from research that is far from immune to criticism.

REFERENCES

- Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, 32, 505-23.
- Gonzalez, E. J., & Kennedy, A. M. (Eds.). (2003). *PIRLS 2001 user guide for the international database*. Boston, MA: Lynch School of Education, Boston College.

- Hattie, J. (2003, October). *Teachers make a difference: What is the research evidence?* Paper presented at the Australian Council for Educational Research Conference.
- IEA. (2003). *PIRLS 2001 user guide for the international database, supplement one, international version of the background questionnaire*. Boston, MA: Lynch School of Education, Boston College.
- It's the teachers, stupid. (2004, June 13). *Sunday Star Times*, p.C10.
- Larkin, S. (2002). Creating metacognitive experiences for 5- and 6-year-old children. In P. Adey & M. Shayter (Eds.), *Learning intelligence: Cognitive acceleration across the curriculum from 5 to 15 years* (pp.65-79). Buckingham: Open University Press.
- Mallard, T. (2003, May 7). Best teaching practices to be developed. Electronic newsletter from announce@executive.govt.nz
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004, April). *How large are teacher effects?* Paper presented at the American Educational Research Association conference, San Diego.
- Timperley, H. S., Phillips, G., & Wiseman, J. (2003). *The sustainability of professional development in literacy. Part I: Changing and sustaining teachers' expectations through professional development in literacy* [Report to Ministry of Education]. Wellington: Ministry of Education, Research Division.
- Timperley, H. S., & Wiseman, J. (2003). *The sustainability of professional development in literacy. Part II: Changing and sustaining teachers' expectations through professional development in literacy* [Report to Ministry of Education]. Wellington: Ministry of Education, Research Division.

Submitted for publication May 27, 2004
Final revision received August 18, 2004
Accepted August 23, 2004



About the Author(s)

New Zealand Journal of Teachers' Work, Volume 1, Issue 1, 2004

ROY NASH

Massey University

Roy Nash is a professor of education at Massey University College of Education. He is a sociologist of education with a speciality in the explanation of social disparities in educational achievement. Recent papers can be found in *New Zealand Journal of Educational Studies*, *Journal of Curriculum Studies*, *Waikato Journal of Education* and *International Studies in Sociology of Education*.

Professor Roy Nash

Department of Social and Policy Studies in Education

Massey University College of Education

Private Bag 11 222

Palmerston North

R.Nash@xtra.co.nz