

# Commentary

## **Invisible Learnings? A Commentary on John Hattie's book: Visible Learning: A synthesis of over 800 meta-analyses relating to achievement**

IVAN SNOOK, JOHN O'NEILL, JOHN CLARK, ANNE-MARIE O'NEILL AND ROGER OPENSHAW

*Massey University*

Evidence does not supply us with rules for action but only with hypotheses for intelligent problem solving, and for making inquiries about our ends in education. (John Dewey, quoted in Hattie, 2008, p.147)

### INTRODUCTION

This book by Professor John Hattie of Auckland University is the result of decades of careful research. He has synthesised some 800 meta-analyses comprising more than 50,000 studies and involving some 146,000 'effect sizes'. The announcement of the book has already led to a good deal of discussion both in New Zealand and overseas and seems to have captured the attention of policy-makers. It is, therefore, important that members of the educational research community pay John Hattie the courtesy of subjecting his conclusions to critical scrutiny in a spirit of mutual truth-seeking to ensure that: (1) discussions are based on a careful reading of the book, rather than on half-baked 'reactions' in the popular media; (2) the caveats which Hattie himself sets out are carefully noted so that decisions are not made in opposition to the message of this book, and (3) the findings are not 'appropriated' by political and ideological interests and used in ways which the data do not substantiate.

### THE METHODOLOGY UNDERLYING THE BOOK

Hattie derives his results from working on a large sample of research studies. His method involves a synthesis of a large number of meta-analyses of studies about education variables. A meta-analysis is a statistical technique for amalgamating, summarising and reviewing primary research. It combines the results of various studies which address a set of research hypotheses. It is used in many branches of knowledge such as medicine, psychotherapy, business and education. All the findings in this book derive from John Hattie's synthesis of 800 meta-analyses of more than 50,000 quantitative studies of variables affecting the achievement of students.

A major aim is to determine effect sizes. From looking at a large number of research studies it is relatively easy to determine that there are certain effects: for example, overall,

drug A is more successful in lowering blood pressure than drug B. But the key question is, “How much more successful?” Effect size is a way of answering this question. It involves comparing the mean scores of the two variables and dividing them by the standard deviation (Coe & Rowe, 2004). Thus, studies can be plotted along a continuum from very low effect size to very high effect size. In both cases a judgement is needed, for although it is not disputed that an effect size of 1.0 is large, there are debates about where a small effect size ends and a moderate or large effect size begins. Hattie adopts 0.4 as the cut-off point, basically ignoring effect sizes lower than 0.4. Thus, for example, class size is interpreted as a small effect size since it is 0.2 (in public debate this tends to turn into “class size has no effect at all”). Selecting a cut-off point is a hazardous exercise, as it means that potentially important effects may be overlooked. An effect size of 0.2 means that the difference between the two comparison groups (e.g. small classes and large classes) is 0.2 (20%) of a standard deviation of the test or measurement scores. Much depends, therefore, on the quality of the research studies in the various meta-analyses. If the sample is large and random (hence increasing the validity and reliability of the measurement), a ‘small’ effect size is of considerable significance. On the other hand, large effect sizes from small samples are meaningless at best and positively dangerous when lumped together with other studies to produce an ‘average’.

Hattie states that he has synthesized 800 meta-analyses, and insists his study is not a meta-analysis of meta-analyses. What is a synthesis? According to the Evidence Informed Policy Network (undated), the term ‘research synthesis’ is defined as a “... systematic and transparent summary of the best available evidence relevant to a policy decision”. The key point is that a synthesis must “... include the development of a protocol, the use of systematic and explicit methods, data collection, analysis, interpretation and reporting of the results.” Hattie says that he is not concerned with the quality of the research in the 800 studies but, of course, quality is everything. Any meta-analysis that does not exclude poor or inadequate studies is misleading, and potentially damaging if it leads to ill-advised policy developments.

Just as this commentary was being finalised, the Ministry of Education and NZCER released an excellent article on effect sizes. It repeats many of the reservations which we express in our commentary (Schagen & Hodgen, 2009). As well, Hattie himself acknowledges many of the problems outlined in this article.

## QUALIFICATIONS OF HIS STUDY

John Hattie himself acknowledges some of the problems associated with his approach:

### **Social effects/background/context effects are ruled out**

[This] is not a book about what cannot be influenced in schools – thus critical discussions about class, poverty, resources in families, health in families, and nutrition are not included but this is NOT because they are unimportant, indeed they may be more important than many of the issues

discussed in this book. It is just that I have not included these topics in my orbit (Hattie, 2008, pp.x-xi).

As we shall see, social class background is indeed more important than many of the issues discussed in this book and hence policy decisions cannot be drawn in isolation from the background variables of class, poverty, health in families and nutrition.

### **The various studies have not been appraised for their validity**

[This] is not a book about criticism of research and I have deliberately not included much about moderators of research findings based on research attributes (quality of study, nature of design) not because these are unimportant ... but because they have been dealt with elsewhere (p.ix).

However, he is not entirely consistent on this. In his discussion of extra-curricular activities, he cautions against taking the finding (0.47) too seriously since it is based on “a random effects” model, which may lead to inflated effect sizes. In relation to charter schools vs regular schools, he cites a study which reports an effect size of 0.2, “... but when the lower quality studies were excluded, this difference dropped to zero” (p.66). In his treatment of ‘learning styles’ he is justifiably suspicious of the motives behind much of the research and appropriately sceptical of many of the results. Thus, although he finds an effect size of 0.41 overall, he dismisses it as not credible (pp.195-197). Once again, it is Hattie’s right to define how he will approach the data but his own approach acknowledges that a simple effect size analysis cannot be maintained, when there are questions about the underlying quality of some of the studies he includes.

### **The research is limited to one dimension of schooling**

Of course there are many outcomes of schooling such as attitudes, physical outcomes, citizenship, and a love of learning. This book focuses on student achievement and that is a limitation of this review. (p.6)

To be more accurate, he is concerned not with achievement but with achievement that is amenable to quantitative measurement. New knowledge, skills and dispositions are all ‘achievements’ of one form or another but they are generally more difficult to measure. At times, his restricted scope leads to arbitrary conclusions. Writing about the effects of programmes of moral education, he says: “The major outcome from moral education programmes is the facilitation of moral judgement ... and as this is not strictly achievement as typically defined, these are not included in the tables” (p.149). He also has to concede that the form of ‘learning’ which he discusses is, itself, severely limited. Having distinguished three levels of learning (surface, deep and conceptual), he says in one of his conclusions: “A limitation of many of the results in this book is that they are more related to the surface and deep knowing and less to conceptual understanding” (p.249). And yet, conceptual knowing or understanding is what he thinks should be the result of good teaching. Clearly there is less to be drawn from his synthesis than commentators have suggested. Much depends on the kind of learning that is desired in formal education. Policy-makers have

to take a broad view of schooling; they have to be interested not just in achievement on narrow tests or even on deeper conceptual knowledge, important as this clearly is, but on the attitudes which students bring to their lives as workers and citizens. Employers, for example, often stress the importance of the attitudes which young people bring to work – perseverance, flexibility, cooperation – rather than only the cognitive qualities that they can demonstrate.

### **The research may not be applicable to ordinary teachers**

Most of the successful effects come from innovations and these effects from innovations may not be the same as the effects of teachers in regular classrooms ... (p. 6)

This is particularly telling when, as in the case of the *Picking up the Pace* studies, we are told that the class size was kept artificially low for the duration of the study (Ministry of Education, undated).

### **Correlation must not be confused with causation**

He also has a very interesting discussion on the importance of not confusing correlation with causation and moving too readily from “this is significant statistically” to “this is what teachers should do” (pp.3-4). As an example, after finding that “feedback” is important, Hattie adds: “It would be an incorrect understanding of the power of feedback if a teacher were to encourage students to provide more feedback” (p.4). He concedes, though, that “... the fundamental word in meta-analysis, effect size, implies causation (what is the effect of a on b) and this claim is often not defensible” (p.237).

## **PROBLEMS WITH THE USE OF META-ANALYSIS**

Hattie has set out some of the major problems with the methodology that he has used for his study. First, comparing disparate studies can be like comparing “apples and oranges.” Each study can be very different. Second, in seeking “averages”, studies ignore the complexity of classrooms and the wide variety of results. Third, what is so sacred about an average score? Fourth, the studies are “historical”; i.e. they report past findings and cannot show that the future must be the same. Fifth, they do not distinguish the quality of different studies and hence could merit Eysenk’s judgment: “garbage in garbage out”. Hattie works to minimise these criticisms of his methodology but they need to be taken into account before accepting the analyses as sound for policy purposes.

There are some other problems (not centrally acknowledged by Hattie) associated with meta-analyses:

- (i) Bias is not normally controlled in meta-analyses. Thus a meta-analysis (however well designed) of poorly designed studies will inevitably lead to unreliable conclusions. It is a serious matter when government agencies use such conclusions to justify educational policy.

(ii) There is a heavy reliance on published results. As we know, particularly in relation to studies commissioned by drug companies (but also from studies of lucrative educational fads such as 'learning styles'), this often means that studies which fail to support favoured conclusions do not make it into publications and thus into meta-analyses. Once again, this has important ramifications for policy-making.

(iii) There is a particular problem in relation to education: the difficulty of clearly defining the variables. In medicine, for example, drug A can be carefully compared to drug B in terms of their respective chemical qualities but it is not nearly so easy when one is talking about such things as child-centred teaching vs teacher-centred teaching. There is no clear operational definition of either of the variables. In these matters there is usually a continuum and, therefore, subjective judgments have to be made: where is the line to be drawn? On one occasion, at least, Hattie himself draws attention to this problem. Writing about the effects of whole language teaching in reading he notes discrepant results from two meta-analyses in which "... there was much overlap in the studies used ... and the difference is a function of how the authors classified some key studies, and the coding of what constituted whole language" (p.137). We suspect that this category of problem might be widespread.

(iv) There is also the difficulty which arises from amalgamating a large number of disparate studies. When results of many studies are averaged, the complexity of education is ignored: variables such as age, ability, gender and subject studied are set aside. An example of this problem can be seen in Hattie's treatment of homework: does homework improve learning or not? Overall, Hattie finds that the effect size of homework is 0.29. Thus a media commentator reading a summary might justifiably report: "Hattie finds that Homework does not make a difference." When, however, we turn to the section on homework we find that, for example, the effect sizes for elementary (primary in our terms) and high schools students are 0.15 and 0.64 respectively. Thus the figures suggest that homework is very important for high school students but relatively unimportant for primary school students. There were also significant differences in the effects of homework in mathematics (high effects) and science and social studies (both low effects). Results were high for low-ability students and low for high-ability students. The nature of the homework set was also influential (pp.234-236). All these complexities are lost in an average effect size of 0.29.

(v) There is also the issue of how generalisable the results are. Hattie points out that most of the studies were carried out in highly developed English-speaking countries (mainly the USA) and should not be generalised to non-English speaking or developing countries. It has been shown, for example, that in developing countries, school effects (as against teacher effects) are huge, due no doubt to the wide variety of schools. It could easily be that New Zealand schools, teachers, students, schools

and curricula are sufficiently different from those of the USA that the meta-analyses may not fully apply in this country.

## SCHOOL EFFECTS

Hattie acknowledges the important role of socio-economic status and home background but chooses to ignore it. That is his choice; but it is easy for those seeking to make policy decisions to forget this significant qualification. There is some debate about the extent of the contribution made by a student's social background but the following conclusions are typical:

(i) Gray, Jesson and Jones (1986) summarised their large-scale research in Britain: "Around 80% of the difference can be explained by the intake" and they say that "... this has held up over all the schools and LEAs studied." They went on to say that half the remaining difference (the 20%) may be explained by the school's examination policies. This would leave only 10% to be explained by other variables within the school.

(ii) Based on his research in New Zealand (and consistent with many overseas studies) Richard Harker has claimed that "... anywhere between 70-80% of the between schools variance is due to the student 'mix' which means that only between 20% and 30% is attributable to the schools themselves" (including, of course, the teachers) (Harker, 1995, p.74). Certainly, he found quite significant differences between schools in their results even after the influence of social background is controlled (the 'value added' effect) (Harker, 1996).

(iii) According to a recent OECD volume on the importance of quality teaching, it is possible to draw three "broad conclusions" from the research on student learning.

The first and most solidly based finding is that the largest source of variation in student learning is attributable to differences in what students bring to school – their abilities and attitudes, and family and community. Such factors are difficult for policy-makers to influence, at least in the short run. (OECD, 2005, p.2).

(iv) Hattie in fact seems to acknowledge this. Although he does not discuss social background he refers to student influences on learning and home influences on learning. In another publication he ascribes 50% of the variance to what the student brings and 10% to the contribution of the home (Hattie, 2003, pp.1-2). Of course, under 'student influences' he includes IQ but seems to see this as a fixed (inherited?) quality rather than the largely socially determined one it is now known to be (Nash, 2004). This leaves only 40% to be explained by school and teacher influences. This is, admittedly, rather larger than most other estimates, but still much smaller than the influence of social background on achievement.

There are; in fact, two different types of research on 'school effects'. One compares the relative contribution made by social variables on the one hand and school variables on the other. The former includes social status, parental education, home resources and the like; the latter includes all variables within the school: curriculum, principal, buildings and the work of teachers. These studies typically find that most of the variance comes from the social variables and only a small part from the school (including the teachers).

The other kind of study is that which ignores the social variables and asks simply: which of the school variables are most important: policies, principal, buildings, school size, curriculum or teachers? These, unsurprisingly, tend to find that the teacher is the most important variable, that is, more important than the principal, the curriculum, the school size or the policies. It is easy to get these two types of studies confused. A former Minister of Education, badly advised by his Ministry, made a fool of himself for some months before making the necessary qualification: that in saying "Teachers are the most important variables in student learning" he was talking about studies of the second type and after being publicly criticised he began to add the crucial qualifier, "within the school". Sadly, in our contemporary politicised and uncritical social climate neither his egregious error nor his retraction was noted by the media, the Ministry, or, by and large, academic commentators.

## OTHER ISSUES

### **Isolating the variables to be analysed**

There is a particular difficulty, noted above, in specifying educational differences for research purposes. For example, with small vs large classes: how does one define 'small' and 'large'? Similarly, with open vs traditional classes: how to estimate the extent of openness etc? Equally, with streamed (tracked) vs unstreamed schools or classes: how much 'streaming' or 'selective grouping' etc is acceptable while the class is still classified as unstreamed? Comparing such abstract variables is not at all like comparing drug A with drug B in medical research or even urban vs rural differences in sociology. Classrooms are very complex and relevant variables are hard to pin down. A problem for meta-analyses is that each study tends to define these factors differently, leading to findings being based on different research conditions.

### **Interpretation of 'small' vs 'big' differences**

Hattie adopts (arbitrarily) a cut-off at 0.4 and above, but other researchers are content with a lower cut-off point. To some extent the choice is arbitrary but, as we said earlier, what is important is not the effect size per se but the quality of the research underlying the meta-analyses. This is what should make the difference when suggestions are made for policy. In fact, Hattie concedes that in some areas a much lower threshold can be significant. In medicine, it was demonstrated that the effect size of taking one low-dose aspirin to decrease the risk of heart attack was a mere 0.07 but it translates into the conclusion that 34 out of 1,000 people would be saved from heart attack. "This sounds worth it to me," he says (p.9). Indeed, Hattie is not always thoroughly consistent in relying on an effect size of at least 0.4. Writing of the studies of outdoor education he finds it "most exciting" that the

“follow up” effects were (untypically) “positive.” The effect size was 0.17, well below his usual cut-off point (p.157). Why is he so excited by this rather modest result when effect sizes higher than this are often written-off as insignificant?

## TWO PARTICULAR ISSUES

### **Class size**

Hattie has been cited as “finding” that class size is not important and this has excited the attention of those concerned about financing of schools, who conclude that they can economise on class size. In fact, the significance of class size is much more complicated than that, even in terms of John Hattie’s synthesis. What is a small class: five, 15, 20? What is a large class: 25, 40, 80? (really large classes are common in tertiary education). It is interesting to note that in the STAR studies (discussed below) classes of 22 to 25 were defined as large, when in many studies these would be seen to be desirably small compared to, say, classes of 30+. It is also important to determine how the assessment is made: on the basis of teacher-pupil ratios in a whole school? (this is quite common, hence we do not know how large any actual class is); on average attendance over a period of time?; or an actual count on the days the teaching is done and the testing carried out? (this would seem to be the most desirable method). Studies vary greatly in relation to these ways of estimating ‘class-size’. By its nature, it is very difficult for meta-analyses to effectively take such differences into account. For Hattie, using meta-analyses as his source material, the problem is further compounded by the distance from original studies.

Hattie concludes that the effect size for class size is around 0.2, which is in his category a small effect. On this basis he seems to dismiss it and commentators in the popular media have played this up. However, some points can be made. First, this is not negligible. Second, many studies have suggested a much higher rating for class size. Prominent among these is the well-known Student/Teacher Achievement Ratio Study (STAR) study.

STAR was set up as a result of some inconclusive debate about class size. Smith and Glass (1980) did a meta-analysis of studies on class size and concluded that “... well-designed studies produced quite different results from studies with minimal controls” (p.429). Adopting stricter criteria they found that small classes have a decided advantage in relation to the attitudes of students (0.47) and teachers (1.03) (a massive effect size in Hattie’s terms, although, of course, he explicitly excludes attitudinal variables from his synthesis) and also in relation to test performance in reading (0.30) and maths (0.32) (Hattie reports lower effect sizes from this study). However, these findings were challenged and the STAR project was set up to try to resolve the impasse. It studied 76 elementary schools in Tennessee in a randomised experiment. ‘Small’ was defined as 13 to 17, ‘large’ as 22 to 25 students. Teachers and students were randomised into small and big classes. The study of achievement was carried out after two years when 6,750 children were subjected to standardised tests of reading and maths on a pass/fail basis where 80% was a pass. Effects sizes varied but there were some at 0.64, 0.66, and 0.62 which are clearly well above Hattie’s cut-off for significance (0.4) and about the same as most of the variables which he regards as very important (Finn & Achilles, 1990). They claim that “... there was a clear positive effect”, particularly for minority groups and particularly in the early years.

Similarly, in Britain, Blatchford and others came to the conclusion that previous studies lacked the design features which would enable sound conclusions to be drawn and they set up 'The Institute of London Class Size Study'. They drew their sample from eight LEAs, 199 schools, 330 classrooms and 7,142 students. They found many positive results for various process and affective aspects of smaller classes and, in relation to attainment which is the focus of the Hattie study, they found that, "There is clear effect of class size on children's academic attainment over the Reception year and there is a clear case for small class sizes during the first year of schooling for both literacy and numeracy" (Blatchford, 2003, p.164). The superior results for literacy were particularly obvious for lower-ability children. While the effects on individuals tended to continue into the second year, the researchers found no clear evidence of class size differences beyond Year 1. Their data provide another cautionary tale: in comparing classes of 15 with classes of 23, large differences were found but there were only negligible differences between classes of, say, 20 and 25 – sometimes in favour of the larger class. This again indicates that 'small' and 'large' are not clearly defined terms and one must constantly be aware of what a particular researcher is studying.

Hattie notes (2008, p.86) that the low effects score for class size may be due to the fact that teachers of smaller classes do not always vary their teaching to take advantage of the smaller group. This is important. Simply reducing class size does nothing to the teaching-learning process. Only if changes are also made to the teaching-learning interaction are any achievement effects possible. This point was demonstrated by Murnane and Levy (1996) who looked at the effects of additional resourcing (USD\$300,000 per annum per school for five years) in a sample of 15 extremely poorly performing (as measured on mandatory state-wide achievement tests) Texas primary schools serving low-income, minority group children. Thirteen of the 15 schools showed no significant changes in student achievement over the course of the study. In these schools, the additional resourcing was used primarily to reduce class size by hiring additional teachers. This result is consistent with Hattie's view that reducing classes makes comparatively little difference to achievement. The other two schools also used much of the money to reduce class sizes, but they also did other things: the principal worked with parents and teachers to confront the problem of low achievement; children with special needs were included in regular (now smaller) classes; teachers' pedagogies were changed by introducing reading and mathematics programmes previously only provided to gifted and talented children in the district; health service provision was brought into the schools; and parents became heavily involved in school governance. After five years, attendance at these two schools was among the highest in the city and test scores had risen to the city average. In terms of accurately analysing the relationship between resources (including smaller classes) and achievement, the study authors make three key points.

First, if the analysis of estimated effects had been conducted after only one year, the data would have shown no effects because the changes in these two schools took several years to take effect. Second, if estimated in conjunction with the data across all 15 schools, the analysis would have shown a small negative relationship to achievement (the average of large effects in two schools and no effects in 13 schools). Third, and most significantly,

they argued that if a model were devised that "... included interactions between class size, instructional techniques, and investments in raising student attendance and increasing parental involvement, the results would show that the package of changes had enormous effects. In contrast, lowering class size and not changing anything else, especially not changing instructional techniques, had no effect on achievement" (Murnane & Levy, 1996, p.95).

How would this study have been categorised by Hattie and where would it sit in his table of intervention effects? Was this Texas case a study of reduced 'class size', changed 'instructional techniques', 'full-service schools', 'parent governance' or something else? One also has to assume that the class size studies Hattie reviewed did not all have the identical 0.2 effect size (it was an average), nor did they all have identical conditions (they did not all replicate the one study design). In other words, even in a study of what he chose to classify as 'class size', other 'confounding' variables would necessarily have been at play, which would also have had an impact on achievement. Hattie recognises that 'class size' cannot usefully be considered in isolation from other potentially important, pedagogically-related variables. Reducing class size may have only a small effect when considered in isolation but that's not the issue. What matters is that reducing class size permits the teacher (and children) to do things differently.

This is acknowledged by the Ministry of Education (undated) when commenting on the PACE research:

The project findings point to a significant relationship between class sizes for new entrants and the gains made in their achievement levels ... For maximum benefit from this kind of approach, it is recommended that class sizes for children in their first year of schooling in low decile schools should not exceed 18 ... The study showed that while class size did make a difference, the smaller the classes the better the outcomes, but only in conjunction with professional development. Without professional development, class size may make no difference.

Interestingly, the issue of class size was emphasised by the co-principal of one of the schools in the PACE study:

The success of the programme has also been attributed to the board of trustees' decision last year to reduce junior class size from 28 students to 15. This has had an amazing impact because the programme has to be done with groups of three children. When you're involved with each group for 10-15 minutes at a time you can't have large numbers in the classroom unless you have the support of a teacher aide. Smaller numbers mean teachers are able to interact a lot easier with the groups and on a more regular basis (Stewart, 2001, unpaginated).

The claimed successes of the PACE programme have been ascribed to innovative teaching techniques but could just as easily be ascribed to the smaller classes, or more likely, to the interaction between the variables.

The point of mentioning these studies is not to argue that Hattie's analysis is wrong, but to indicate that drawing policy conclusions about the unimportance of class size may be unwarranted and possibly very damaging to the education of children, particularly young children and lower-ability children. A much wider and in-depth analysis is needed.

## **Performance pay**

Hattie's conclusions about the importance of what teachers do has led some to advocate performance pay (sometimes, in the past, called 'merit pay' or 'payment by results'). There have been many attempts at instituting this, particularly in the USA. The judgement of a group of researchers some years ago still stands: "The promise of merit pay is dimmed by knowledge of its history; most attempts to implement merit pay for public school teachers over the past twenty-five years have failed" (Murnane & Cohen, 1981).

The idea has been mooted in New Zealand. In 1985-86 a parliamentary select committee produced the excellent *Report on the Enquiry into the Quality of Teaching (The Scott Report)* (Education and Science Select Committee, 1986). Among the five members of this committee was Ruth Richardson, who was campaigning for a voucher system of education. She would be a 'dry' Minister of Finance in the National Government after 1990 and was certainly no 'bleeding heart liberal' or 'lackey of the teacher unions'. As was to be expected from its composition, the committee produced a hard-hitting report which argued that measures of teacher performance were urgently needed but acknowledged that the process of developing such measures will be "lengthy and complex" and it advocated the setting up of a research unit based at a university to try to develop sound measures. No such group has ever been set up and no such measures have been developed for New Zealand schools. This might suggest that rushing into a scheme in the 21<sup>st</sup> century would not be a smart idea, particularly as the public are rightly shocked at seeing huge ('performance'!) payouts to managers whose enterprises have failed.

In the USA in particular there have been many more attempts to institute performance pay over the past 25 years and there are varying reports of their successes and failures. However, we have seen no evidence at all to support the claim that performance pay improves teaching or learning and there is nothing in Hattie's massive research which even remotely suggests that it does. On the contrary, much of what he says suggests the very opposite. He says, for example:

School leaders and teachers need to create school, staffroom, and classroom environments where error is welcome as a learning opportunity, where discarding incorrect knowledge and understandings is welcomed, and where participants can feel safe to learn, re-learn, and explore knowledge and understanding. (p.239)

He goes on to add that what is needed for school improvement is "... a caring, supportive staff room, a tolerance for errors, and for learning from other teachers, a peer culture among teachers of engagement, trust, shared passion, and so on" (p.240). Such a co-operative, trusting and self-critical school atmosphere is the very kind of atmosphere which regimes of performance pay destroy.

## SIGNIFICANCE FOR POLICY AND PRACTICE

Teachers must learn to take account of research findings even when (particularly when) they go against long-held beliefs. Hattie draws attention to a situation (p.258) where teachers ignored evidence in favour of their own deeply-held beliefs. Teaching will never make progress as a profession while this unwillingness persists.

However, the following comment of the late Roy Nash (whose contribution to debates on these topics was unequalled and is deeply missed) is apposite:

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 (Nash, 2004, p.49).

Policy-makers must learn that research data cannot be automatically applied to practice. Knowing, for example, that teachers should establish good feedback arrangements with pupils does not tell any teacher what she is to do. Research knowledge has to be synthesised and integrated with the teacher's beliefs, values and experience. Hattie fully acknowledges this, following Dewey in holding that: "Evidence does not supply us with rules for action but only with hypotheses for intelligent problem solving, and for making inquiries about our ends in education" (p.247). There is also an irreducible value component to every teaching decision: is the benefit of X sufficient to justify the cost (in terms of money and energy) of instituting it? The presumed benefit must also be weighed against possible or proven harm, for example, attainment on a test might be improved a little by methods which inhibit the creativity of students or damage their ability to relate to others.

Teacher educators must resist the temptation to simplify research evidence for students under facile claims that "... research has shown ...". Unfortunately, the kind of conclusions presented in this book readily lend themselves to such treatment even though Hattie explicitly warns against using his material in this way.

## CONCLUSION

In conclusion, we want to repeat our belief that John Hattie's book makes a significant contribution to understanding the variables surrounding successful teaching and think that it is a very useful resource for teacher education. We are concerned, however, that:

- (i) Despite his own frequent warnings, politicians may use his work to justify policies which he does not endorse and his research does not sanction;
- (ii) Teachers and teacher educators might try to use the findings in a simplistic way and not, as Hattie wants, as a source for "hypotheses for intelligent problem solving";
- (iii) The quantitative research on 'school effects' might be presented in isolation from their historical, cultural and social contexts, and their interaction with home and community backgrounds; and

(iv) In concentrating on measureable school effects there may be insufficient discussion about the aims of education and the purposes of schooling without which the studies have little point.

It is important that students preparing for teaching learn about the research process and how easily it leads to error rather than truth. They need to respect research but be acutely aware of its limitations. The research that they need to know about goes beyond what happens in schools and classrooms. The things that students bring from their social class, family, culture, home background and prior experiences are more important than what happens in the school, even though what happens in the school (particularly what teachers are and do) is very important. The secret of school improvement lies in the recognition of all these factors and their integration into a social, economic and educational programme.

## REFERENCES

- Blatchford, P. (2003). *The class size debate: Is small better?* Maidenhead, UK: Open University Press.
- Burtless, G. (Ed). (1996). *Does money matter? The effect of school resources on student achievement and adult success* (pp. 93-96). Washington: Brookings Institution Press.
- Coe, R. & Rowe, K. (2004). *What is an 'effect size'?* Camberwell, VIC, Australia: Australian Council for Educational Research.
- Education and Science Select Committee. (1986). *Report on the enquiry into the quality of teaching (The Scott Report)*. Wellington, NZ: Government Printer.
- Evidenced Informed Policy Network. (Undated). *Policy Synthesis*. <http://www.evipnet.org/php/level.php?lang=en&component=101&item=2> (retrieved 13/2/2009).
- Finn, J.D. & Achilles, C.M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27(3), 557-577.
- Gray, J., Jesson, D. & Jones, B. (1986). Towards a framework for interpreting examination results. In R. Rodgers (Ed). *Education and social class* (pp.51-57). London, UK: Falmer Press.
- Harker, R. (1995). Further comment on 'So Schools Matter?'. *New Zealand Journal of Educational Studies*, 30(1), 73-76.
- Harker, R. (1996). On 'First year university performance as a function of type of secondary school attended and gender'. *New Zealand Journal of Educational Studies*, 32(2), 197-198.
- Hattie, J. (2003). *Teachers make a difference: What is the research evidence?* Paper presented to the Australian Council for Educational Research Annual Conference 'Building Teacher Quality' held in October 2003. <http://www.leadspace.govt.nz/leadership/articles/teachers-make-a-difference.php> (retrieved 13/2/2009).
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London. UK: Routledge.

- Ministry of Education. (Undated). *Picking up the Pace*. <http://www.minedu.govt.nz/educationSectors/PasifikaEducation/ResearchAndStatistics/PickingUpThePace.aspx> (retrieved 16/2/2009).
- Murnane, R. & Cohen D. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Education Review*, 56(1), 1-17.
- Nash, R. (2004). Teacher effects and the explanation of social disparities. *New Zealand Journal of Teachers Work*, 1(1), 42-50.
- OECD. (2005). *Teachers matter: Attracting, developing and retaining effective teachers. Overview*. Paris, France: OECD. <http://www.oecd.org/dataoecd/39/47/34990905.pdf> (retrieved 13/2/2009).
- Schagen, I. & Hodgen, E. (2009). *How much difference does it make? Notes on understanding, using and calculating effects sizes for schools*. <http://www.educationcounts.govt.nz/publications/schooling/36098> (retrieved 30/3/2009).
- Smith M.L. & Glass, G.V. (1980). Meta-analyses of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal*, 17, 419-433.
- Stewart, K. (2001). Elated but not sated. *Education Gazette*, 80(22). <http://www.fulbright.org.nz/fulbrighthays-2003/projects/resources.html> (retrieved 13/2/2009).

Manuscript received: 7 April 2009

Revisions received: 26 May 2009

Accepted: 7 June 2009

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.