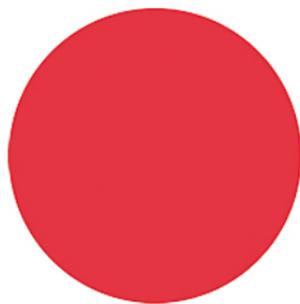
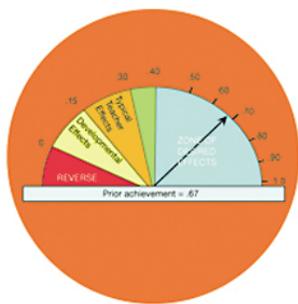
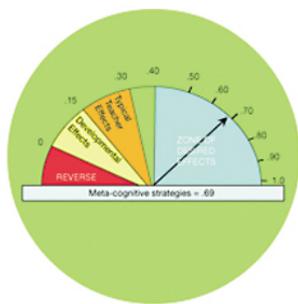
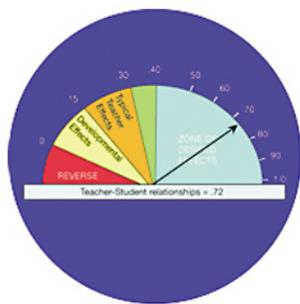
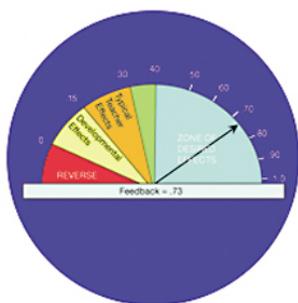
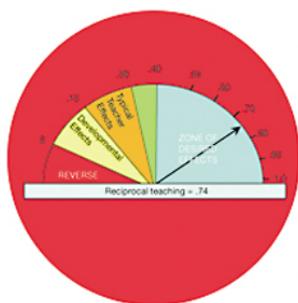
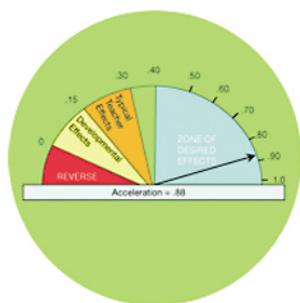


# VISIBLE LEARNING

## A SYNTHESIS OF OVER 800 META-ANALYSES RELATING TO ACHIEVEMENT



JOHN HATTIE



---

# Visible Learning

---

This unique and ground-breaking book is the result of 15 years' research and synthesises over 800 meta-analyses relating to the influences on achievement in school-aged students. It builds a story about the power of teachers and of feedback, and constructs a model of learning and understanding.

*Visible Learning* presents research involving many millions of students and represents the largest ever collection of evidence-based research into what actually works in schools to improve learning. Areas covered include the influences of the student, home, school, curricula, teacher, and teaching strategies. A model of teaching and learning is developed based on the notion of visible teaching and visible learning.

A major message within the book is that what works best for students is similar to what works best for teachers. This includes an attention to setting challenging learning intentions, being clear about what success means, and an attention to learning strategies for developing conceptual understanding about what teachers and students know and understand.

Although the current evidence-based fad has turned into a debate about test scores, this book is about using evidence to build and defend a model of teaching and learning. A major contribution to the field, it is a fascinating benchmark for comparing many innovations in teaching and schools.

**John Hattie** is Professor of Education and Director of the Visible Learning Labs, University of Auckland, New Zealand.



---

# Visible Learning

---

A synthesis of over 800 meta-analyses  
relating to achievement

John A. C. Hattie

First published 2009  
by Routledge  
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada  
by Routledge  
270 Madison Avenue, New York, NY 10016

*Routledge is an imprint of the Taylor & Francis Group, an informa business*

This edition published in the Taylor & Francis e-Library, 2008.

“To purchase your own copy of this or any of Taylor & Francis or Routledge’s collection of thousands of eBooks please go to [www.eBookstore.tandf.co.uk](http://www.eBookstore.tandf.co.uk).”

© 2009 John A. C. Hattie

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

*British Library Cataloguing in Publication Data*

A catalogue record for this book is available from the British Library

*Library of Congress Cataloging-in-Publication Data*

Hattie, John.

Visible learning: a synthesis of meta-analyses relating to achievement/John A. C.

Hattie.

p. cm.

Includes bibliographical references.

1. Learning—Longitudinal studies. 2. Teaching—Longitudinal studies. 3. Effective teaching—Longitudinal studies. 4. Teacher effectiveness—Longitudinal studies.

I. Title.

LB1060.H388 2008

370.15'23—dc22

2008021702

ISBN 0-203-88733-6 Master e-book ISBN

ISBN10: 0-415-47617-8 (hbk)

ISBN10: 0-415-47618-6 (pbk)

ISBN10: 0-203-88733-6 (ebk)

ISBN13: 978-0-415-47617-1 (hbk)

ISBN13: 978-0-415-47618-8 (pbk)

ISBN13: 978-0-203-88733-2 (ebk)

---

# Contents

---

<i>List of tables</i>	vi
<i>List of figures</i>	vii
<i>Preface</i>	viii
<i>Acknowledgments</i>	x
1 The challenge	1
2 The nature of the evidence: a synthesis of meta-analyses	7
3 The argument: visible teaching and visible learning	22
4 The contributions from the student	39
5 The contributions from the home	61
6 The contributions from the school	72
7 The contributions from the teacher	108
8 The contributions from the curricula	129
9 The contributions from teaching approaches—part I	161
10 The contributions from teaching approaches—part II	200
11 Bringing it all together	237
<i>Appendix A: the meta-analyses by topic</i>	263
<i>Appendix B: the meta-analyses by rank order</i>	297
<i>Bibliography</i>	301
<i>Index</i>	375

---

# Tables

---

2.1 Average effect for each of the major contributors to learning	18
4.1 Summary information from the meta-analyses on the contributions from the student	39
5.1 Summary information from the meta-analyses on the contributions from the home	61
6.1 Summary information from the meta-analyses on the contributions from the school	74
6.2 Synthesis of meta-analyses and major studies reducing class size from 25 to 15	87
7.1 Summary information from the meta-analyses on the contributions from the teacher	109
8.1 Summary information from the meta-analyses on the contributions from the curricula	130
9.1 Summary information from the meta-analyses on the contributions from teaching approaches	162
9.2 Relation between goal difficulty and performance	165
9.3 Difficult compared to “do your best” goals	165
9.4 Relation of self-efficacy to goal attainment	166
9.5 Various metacognitive strategies and the effect sizes (Lavery, 2008)	190
10.1 Summary information from the meta-analyses on the contributions from teaching approaches	201
10.2 Effect sizes for various teaching strategies (from Seidel & Shavelson, 2007)	202
10.3 Effect sizes for various teaching strategies (from Marzano 1998)	203
10.4 Summary of effects from comprehensive teaching reforms (Borman <i>et al.</i> , 2003)	216
10.5 Summary of effects from computer-based instruction	222
10.6 Summary of effects from computers as substitute and as supplement to the teacher	223
10.7 Summary of effects from using computers with the same or a different teacher	223
10.8 Summary of major uses of computers in classrooms	224
11.1 Effect sizes for teacher as activator and teacher as facilitator	243
11.2 Effect sizes from teaching or working conditions	244

---

# Figures

---

1.1	Percentage of responses as to the claimed influences on student learning by students, parents, principals, and teachers	5
2.1	An achievement continuum	7
2.2	Distribution of effect sizes across all meta-analyses	16
2.3	Number of meta-analyses above and below the h-point	19
2.4	A typical barometer of influence	19
2.5	Funnel plot of the effect size and sample size from each meta-analysis	21
7.6	Effect sizes for nine teacher–student relationship variables	119
8.10	The percentage of students performing below, at, and above expectation in each year level	141
8.24	Higgins <i>et al.</i> four-part thinking model	156
9.9	A model of feedback	176
10.18	The number of computer-based meta-analyses and their overall effect size	220
10.19	Relation between effect sizes for computer-based instruction and year of publication	221
11.1	A model of Visible teaching – Visible learning	238
11.2	The means for the National Board certified teachers (NBCTs) and non-National Board certified teachers (non-NBCTs), and the effect size of the difference between these two groups	260
11.3	Percentage of student work from NBCTs and non-NBCTs classified as surface or deep learning	260

---

# Preface

---

Elliott is my hero. On his fifth birthday he was diagnosed with leukemia, and this past year has been his *annus horribilis*. On the day of the diagnosis, it was impressive to see the medical team immediately begin interventions. While they aimed to make Elliott stable, the diagnosis regime burst into action. They knew which tests were needed to make the correct diagnosis and when they were satisfied with the initial diagnosis they immediately moved to interventions. Thus began a year of constant monitoring and feedback to the medical team about Elliott's progress. All throughout they collected evidence of progress, they knew what success looked like, and kept all informed about this evidence. Elliott went through many ups and downs, lost his hair (as did I when he gave me a No. 1 cut as his Christmas present, although I drew a line when he asked to shave my eyebrows off as well), and had daily injections in the front of his legs, but he never balked, and throughout the treatment maintained his sparkly personality. The family was never in the dark about what was happening, books were provided, sessions offered, and support for treatment was excellent. The messages in this book owe a lot to Elliott.

This book started in Gil Sax's office in 1990 searching and coding meta-analyses. Motivation to continue the search was inspired by Herb Walberg, and continued in Perth in Australia, North Carolina in the US, and finished here in Auckland in New Zealand. It is a journey that has taken 15 years. The messages have been questioned, labelled provocative, liked, and dismissed, among other more positive reactions. The typical comments are: "the results do not mirror my experience", "why have you not highlighted my pet method", "you are talking about averages and I'm not average", and "you are missing the nuances of what happens in classrooms". There are many criticisms and misunderstandings about what I am *and* am not saying.

So let me start with what this book is not.

- 1 It is *not* a book about classroom life, and does not speak to the nuances and details of what happens within classrooms. Instead it synthesizes research based on what happens in classrooms; as it is more concerned with main effects than interactions. Although I have spent many hundreds of hours in classrooms in many countries, have observed, interviewed, and aimed to dig quite deeply into the nuances of classrooms, this book will not show these details of class living.
- 2 It 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 influences discussed in this book. It is just that I have not included these topics in my orbit.

- 3 It is *not* a book that includes qualitative studies. It only includes studies that have used basic statistics (means, variances, sample sizes). Again, this should not mean qualitative studies are not important or powerful but just that I have had to draw some lines around what can be accomplished over a 15-year writing span.
- 4 It 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) again not because these are unimportant (my expertise is measurement and research design), but because they have been dealt with elsewhere by others (e.g., Lipsey & Wilson, 1993; Sipe & Curlette, 1996a, 1996b).

Rather this is a book about synthesizing many meta-analyses. It is based on over 50,000 studies, and many millions of students—and this is a cut down version of what I could have included as I also collected studies on affective and physical outcomes and on many other outcomes of schooling. I occasionally receive emails expressing disbelief that I have had the time to read so many studies. No, I have not read all primary studies, but as will be seen I have read all meta-analyses, and in some cases many of the primary studies. I am an avid reader, thoroughly enjoy learning the arts of synthesizing and detecting main ideas, and want to create explanations from the myriad of ideas in our discipline. The aim of this book is not to overwhelm with data—indeed my first attempt was discarded after 500 pages of trenchant details; who would care about such details? Instead this book aims to have a message, a story, and a set of supporting accounts of this story.

The message about schools is a positive one. So often when talking about the findings in this book, teachers think I am attacking them as below average, non-thinking, boring drones. In New Zealand, for example, it is clear to me why we rank in the top half-dozen nations in reading, mathematics, and science—we have a nation of excellent teachers. They exist and there are many of them. This book is a story of many real teachers I have met, seen, and some who have taught my own boys. Many teachers already think in the ways I argue in this book; many are seeking to always improve and constantly monitor their performances to make a difference to what they do; and many inspire the love of learning that is one of the major outcomes of any school. This is not a book claiming that teachers are below par, that the profession is terrible, and that we all need to “put in more effort and do better”. Nearly all studies in the book are based on real students in front of real teachers in real schools—and that so many of the effects are powerful is a testament that excellence is happening. *The* major message is that we need a barometer of what works best, and such a barometer can also establish guidelines as to what is excellent—too often we shy from using this word thinking that excellence is unattainable in schools. Excellence is attainable: there are many instances of excellence, some of it fleeting, some of it aplenty. We need better evaluation to acknowledge and esteem it when it occurs—as it does.

---

# Acknowledgments

---

There are so many who have contributed to the data, the book, and the message, and who have provided feedback over these past 15 years: Nola Purdie, Krystoff Krawowski, Richard Fletcher, Thakur Karkee, Earl Irving, Trisha Lundberg, Lorrae Ward, Michael Scriven, Richard Jaeger, Geoff Petty, and Russell Bishop. I am especially indebted to Janet Rivers for her attention to the details and to Debbie Waayer for her remarkable skills in finding articles, referencing, and data skills, and in ensuring that I completed this book. Others have been critics and this is among the more welcome contributions for any author: Lexie Grudnoff, Gavin Brown, Adrienne Alton-Lee, Christine Rubie-Davis, Misty Sato, David Moseley, Heidi Leeson, Brian Marsh, Sandra Frid, Sam Stace, and John Locke. I particular thank Gene Glass for his development of meta-analysis that allowed me and many others to stand on his shoulders to peer into what makes a difference to teaching and learning.

But most of all I thank my family—they have endured this book, shaped the many versions of the message, and provided the feedback that only a loving family can give. Unlike most children who are asked about their day at school each night at the dinner table, my boys have endured the same interrogation every night of their school years: What feedback did you receive about your learning today? Thanks to my boys—Joel, Kyle, Kieran, Billy, Bobby, and Jamie—you are my inspirations for living. And most of all to Janet—the one who has given unconditional positive regard through the ups and downs of moving a family across many countries, putting up with “yet another study”, and being the love of my life. The size of your effect on my life exceeds any reported in this book.

# The challenge

---

In the field of education, one of the most enduring messages is that “everything seems to work”. It is hard to find teachers who say they are “below average” teachers, and everyone (parent, politician, school leader) has a reason why their particular view about teaching or school innovation is likely to be successful. Indeed, rhetoric and game-play about teaching and learning seems to justify “everything goes”. We acknowledge that teachers teach differently from each other; we respect this difference and even enshrine it in terms like “teaching style” and “professional independence”. This often translates as “I’ll leave you alone, if you leave me alone to teach my way.” While teachers talk to their colleagues about curriculum, assessment, children, and lack of time and resources, they rarely talk about their teaching, preferring to believe that they may teach differently (which is acceptable provided they do not question one another’s right to teach in their particular ways). We pass laws that are more about structural concerns than about teaching concerns: such as class size, school choice, and social promotion, as if these are clear winners among the top-ranking influences on student learning. We make school-based decisions about ability grouping, detracking or streaming, and social promotion, again appealing to claims about influences on achievement. For most teachers, however, teaching is a private matter; it occurs behind a closed classroom door, and it is rarely questioned or challenged. We seem to believe that every teacher’s stories about success are sufficient justification for leaving them alone. We will see throughout this book that there is a good reason for acknowledging that most teachers can demonstrate such success. Short of unethical behaviors, and gross incompetence, there is much support for the “everything goes” approach. However herein lies a major problem.

It is the case that we reinvent schooling every year. Despite any successes we may have had with this year’s cohort of students, teachers have to start again next year with a brand new cohort. The greatest change that most students experience is the level of competence of the teacher, as the school and their peers typically are “similar” to what they would have experienced the previous year. It is surely easy to see how it is tempting for teachers to re-do the successes of the previous year, to judge students in terms of last year’s cohort, and to insist on an orderly progression through that which has worked before. It is required of teachers, however, that they re-invent their passion in their teaching; they must identify and accommodate the differences brought with each new cohort of students, react to the learning as it occurs (every moment of learning is different), and treat the current cohort of students as if it is the first time that the teacher has taught a class—as it is for the students with this teacher and this curricula.

As will be argued throughout this book, the act of teaching reaches its epitome of

success after the lesson has been structured, after the content has been delivered, and after the classroom has been organized. The art of teaching, and its major successes, relate to “what happens next”—the manner in which the teacher reacts to how the student interprets, accommodates, rejects, and/or reinvents the content and skills, how the student relates and applies the content to other tasks, and how the student reacts in light of success and failure apropos the content and methods that the teacher has taught. Learning is spontaneous, individualistic, and often earned through effort. It is a timeworn, slow and gradual, fits-and-starts kind of process, which can have a flow of its own, but requires passion, patience, and attention to detail (from the teacher and student).

### So much evidence

The research literature is rich in recommendations as to what teachers and schools should do. Carpenter (2000), for example, counted 361 “good ideas” published in the previous ten years of *Phi Delta Kappan* (e.g., Hunter method, assertive discipline, Goals 2000, TQM, portfolio assessment, essential schools, block scheduling, detracking, character education). He concluded that these good ideas have produced very limited gains, if any. Similarly, Kozol (2005, p. 193) noted that there have been “galaxies of faded names and optimistic claims,” such as “Focus Schools”, “Accelerated Schools”, “Blue Ribbon Schools”, “Exemplary Schools”, “Pilot Schools”, “Model Schools”, “Quality Schools”, “Magnet Schools”, and “Cluster Schools”—all claiming they are better and different, with little evidence of either. The research evidence relating to “what works” is burgeoning, even groaning, under a weight of such “try me” ideas. Most are justified by great stories about lighthouse schools, inspiring principals and inspiring change agents, and tales of wonderful work produced by happy children with contented parents and doting teachers. According to noted change-theory expert, Michael Fullan, one of the most critical problems our schools face is “not resistance to innovation, but the fragmentation, overload, and incoherence resulting from the uncritical and uncoordinated acceptance of too many different innovations (Fullan & Stiegelbauer, 1991, p. 197). Richard Elmore (1996) has long argued that education suffers not so much from an inadequate *supply* of good programs as from a lack of *demand* for good programs—and instead we so often *supply* yet another program rather than nurture *demand* for good programs.

There is so much known about what makes a difference in the classroom. A glance at the journals on the shelves of most libraries, and on web pages, would indicate that the state of knowledge in the discipline of education is healthy. The worldwide picture certainly is one of plenty; we could have a library solely consisting of handbooks about teaching, most of which cannot be held in the hand. Most countries have been through many waves of reform, including new curricula, new methods of accountability, reviews of teacher education, professional development programs, charter schools, vouchers, and management models. We have blamed the parents, the teachers, the classrooms, the resources, the textbooks, the principals, and even the students. Listing all the problems and all the suggested remedies could fill this book many times over.

There are thousands of studies promulgating claims that this method works or that innovation works. We have a rich educational research base, but rarely is it used by teachers, and rarely does it lead to policy changes that affect the nature of teaching. It may be that the research is written in a non-engaging style for teachers, or maybe when research is presented to teachers it is done in a manner that fails to acknowledge

that teachers come to research with strong theories of their own about what works (for them). Further, teachers are often very “context specific”, as the art for many of them is to modify programs to fit their particular students and teaching methods—and this translation is rarely acknowledged.

How can there be so many published articles, so many reports providing directions, so many professional development sessions advocating this or that method, so many parents and politicians inventing new and better answers, while classrooms are hardly different from 200 years ago (Tyack & Cuban, 1995)? Why does this bounty of research have such little impact? One possible reason is the past difficulties associated with summarizing and comparing all the diverse types of evidence about what works in classrooms. In the 1970s there was a major change in the manner that we reviewed the research literature. This approach offered a way to tame the massive amount of research evidence so that it could offer useful information for teachers. The predominant method had always been to write a synthesis of many published studies in the form of an integrated literature review. However in 1976 Gene Glass introduced the notion of meta-analysis—whereby the effects in each study, where appropriate, are converted to a common measure (an effect size), such that the overall effects could be quantified, interpreted, and compared, and the various moderators of this overall effect could be uncovered and followed up in more detail. Chapter 2 will outline this method in more detail. This method soon became popular and by the mid 1980s more than 100 meta-analyses in education were available. This book is based on a synthesis (a method referred to by some as meta-meta-analysis) of more than 800 meta-analyses about influences on learning that have now been completed, including many recent ones. It will develop a method such that the various innovations in these meta-analyses can be ranked from very positive to very negative effects on student achievement. It demonstrates that the reason teachers can so readily convince each other that they are having success with their particular approach is because the reference point in their arguments is misplaced. Most importantly, it aims to derive some underlying principles about why some innovations are more successful than others in influencing student achievement.

### **An explanatory story, not a “what works” recipe**

The aim is to provide more than a litany of “what works”, as too often such lists provide yet another set of recommendations devoid of underlying theory and messages, they tend to not take into account any moderators or the “busy bustling business” of classrooms, and often they appeal to claims about “common sense”. If common sense is the litmus test then everything could be claimed to work, and maybe therein lies the problems with teaching. As Glass (1987) so eloquently argued when the first *What Works: Politics and research* was released, such appeals to common sense can mean that there is no need for more research dollars. Such claims can ignore the realities of classroom life, and they too often mistake correlates for causes. Michael Scriven (1971; 1975; 2002) has long written about mistaking correlates of learning with causes. His claim is that various correlates of school outcomes, say the use of advance organizers, the maintenance of eye contact, or high time on task, should not be confused with good teaching. While these may indeed be correlates of learning, it is still the case that good teaching may include none of these attributes. It may be that increasing these behaviors in some teachers also leads to a decline in other attributes (e.g., caring and respect for students). Correlates, therefore, are not to be confused with the causes.

For example, one of the major results presented in this book relates to increasing the amount of feedback because it is an important correlate of student achievement. However, one should not immediately start providing more feedback and then await the magical increases in achievement. As will be seen below, increasing the amount of feedback in order to have a positive effect on student achievement requires a change in the conception of what it means to be a teacher; it is the feedback to the teacher about what students can and cannot do that is more powerful than feedback to the student, and it necessitates a different way of interacting and respecting students (but more on this later). It would be an incorrect interpretation of the power of feedback if a teacher were to encourage students to provide more feedback. As Nuthall (2007) has shown, 80% of feedback a student receives about his or her work in elementary (primary) school is from other students. But 80% of this student-provided feedback is incorrect! It is important to be concerned about the climate of the classroom before increasing the amount of feedback (to the student or teacher) because it is critical to ensure that “errors” are welcomed, as they are key levers for enhancing learning. It is critical to have appropriately challenging goals as then the amount and directedness of feedback is maximized. Simply applying a recipe (e.g., “providing more feedback”) will not work in our busy, multifaceted, culturally invested, and changing classrooms.

The wars as to what counts as evidence for causation are raging as never before. Some have argued that the only legitimate support for causal claims can come from randomized control trials (RCTs, i.e., trials in which subjects are allocated to an experimental or a control group according to a strictly random procedure). There are few such studies among the many outlined in this book, although it could be claimed that there are many “evidence-informed” arguments in this book. While the use of randomized control trials is a powerful method, Scriven (2005) has argued that a higher gold standard relates to studies that are capable of establishing conclusions “beyond reasonable doubt”. Throughout this book, many correlates will be presented, as most meta-analyses seek such correlates of enhanced student achievement. A major aim is to weave a story from these data that has some convincing power and some coherence, although there is no claim to make these “beyond reasonable doubt”. Providing explanations is sometimes more difficult than identifying causal effects.

Most of these claims about design and RCTs are part of the move towards evidence-based decision making, and the current debate about influences on student learning is dominated by discussion of the need for “evidence”. Evidence-based this and that are the buzz words, but while we collect evidence, teachers go on teaching. The history of teaching over the past 200 years has attested the enduring focus of teachers on notions of “what works” despite the number of solutions urging teachers to move in a different direction. Such “what works” notions rarely have high levels of explanatory power. The model I will present in Chapter 3 may well be speculative, but it aims to provide high levels of explanation for the many influences on student achievement as well as offer a platform to compare these influences in a meaningful way. And while I must emphasize that these ideas are clearly speculative, there is both solace and promise in the following quotation from Popper:

Bold ideas, unjustified anticipations, and speculative thought, are our only means for interpreting nature: our only organon, our only instrument, for grasping her. And we must hazard them to win our prize. Those among us who are unwilling to expose their ideas to the hazard of refutation do not take part in the scientific game.

(Popper, K. R., 1968, p. 280)

## While we collect evidence, teachers go on teaching

As already noted, the practice of teaching has changed little over the past century. The “grammar” of schooling, in Tyack and Cuban’s (1995) terms, has remained constant: the age-grading of students, division of knowledge into separate subjects, and the self-contained classroom with one teacher. Many innovations have been variously “welcomed, improved, deflected, co-opted, modified, and sabotaged” (p. 7), and schools have developed rules and cultures to control the way people behave when in them. Most of us have been “in school” and thus know what a “real school” is and should be. The grammar of schooling has persisted partly because it enables teachers to discharge their duties in a predictable fashion, cope with the everyday tasks that others expect of them, and provide much predictability to all who encounter schools.

One of the “grammars of schooling” is that students are to be made responsible for their learning. This can easily turn into a conception that some students are deficient in their desire for, and achievements from teaching. As Russell Bishop and his colleagues have demonstrated, such deficit thinking is particularly a problem when teachers are involved with minority students (e.g., Bishop, Berryman, & Richardson, 2002). From their interviews, they illustrated that the influences on Māori students’ educational achievement differed for each of parents, students, principals, and teachers (Figure 1.1). Students, parents, and principals see the relationships between teachers and students as having the greatest influence on Māori students’ educational achievement. In contrast, teachers identify the main influences on Māori students’ educational achievement as being Māori students themselves, their homes and/or the structure of the schools. Teachers engage in the discourse of the child and their home by pathologising Māori students’ lived experiences and by explaining their lack of educational

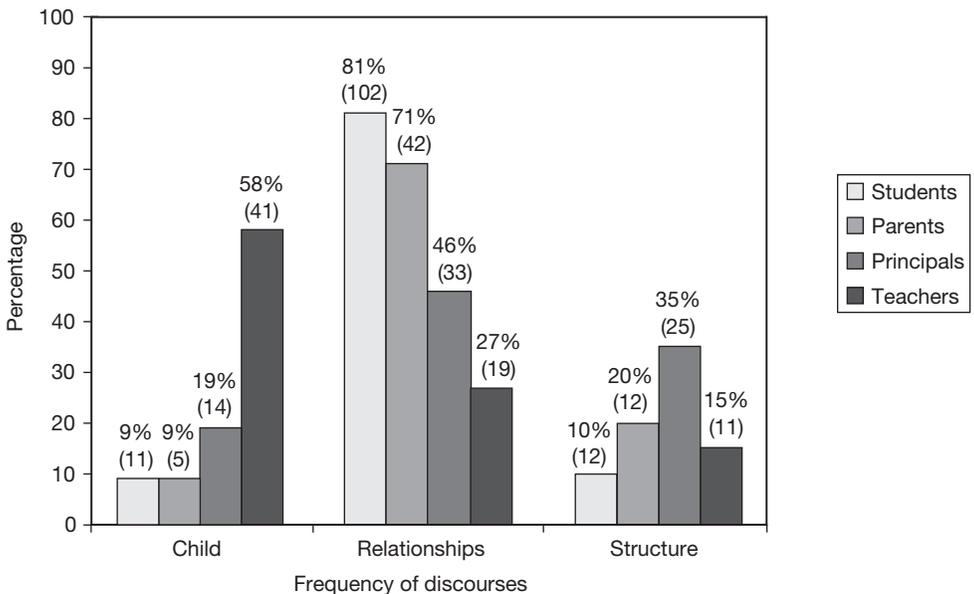


Figure 1.1 Percentage of responses as to the claimed influences on student learning by students, parents, principals, and teachers

achievement in deficit terms. My colleague Alison Jones calls this type of thinking a “discourse of disadvantage” (Jones & Jacka, 1995). They do not see themselves as the agents of influence, see very few solutions, and see very little that they can do to solve the problems.

From their extensive classroom observations, analyses of achievement results, and working with teachers of minority students, Bishop *et al.* have devised a model of teaching Māori students based on caring for all students, and the primacy of the act of teaching. The major features of Bishop’s model include the creation of a visible, appropriate context for learning such that the student’s culture is involved in a process of co-learning, which involves the negotiation of learning contexts and content. The teacher provides supportive feedback and helps students to learn by acknowledging and using the students’ prior knowledge and experiences, and monitoring to check if students know what is being taught, what is to be learnt, or what is to be produced. It involves the teacher teaching the students something, instructing them how to produce something, and giving them instructions as to the processes of learning. This is a high level of teaching activity, indeed.

## **Concluding comments**

This introduction has highlighted the amazing facility of those in the education business to invent solutions and see evidence for their pet theories and for their current actions. Everything seems to work in the improvement of student achievement. There are so many solutions and most have some form of evidence for their continuation. Teachers can thus find some support to justify almost all their actions—even though the variability about what works is enormous. Indeed, we have created a profession based on the principle of “just leave me alone as I have evidence that what I do enhances learning and achievement”.

One aim of this book is to develop an explanatory story about the key influences on student learning—it is certainly not to build another “what works” recipe. The major part of this story relates to the power of directed teaching, enhancing what happens next (through feedback and monitoring) to inform the teacher about the success or failure of their teaching, and to provide a method to evaluate the relative efficacy of different influences that teachers use.

It is important from the start to note at least two critical codicils. Of course, there are many outcomes of schooling, such as attitudes, physical outcomes, belongingness, respect, citizenship, and the love of learning. This book focuses on student achievement, and that is a limitation of this review. Second, 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—the mere involvement in asking questions about the effectiveness of any innovation may lead to an inflation of the effects. This matter will be discussed in more detail in the concluding chapter, where an attempt is made to identify the effects of “typical” teachers compared to “innovations” in teaching. Indeed, the role of “teaching as intervention” is developed throughout the chapters in this book.

# The nature of the evidence

## A synthesis of meta-analyses

---

It is the mark of an educated man ... that in every subject he looks for only so much precision as its nature permits.

(Aristotle, 350BC)

This chapter outlines the methodology relating to the evidence used in the remainder of this book. The fundamental unit of analysis is 800+ meta-analyses and how the major results from these studies can be placed along a single continuum. The chapter then outlines some of the problems of meta-analyses, discusses some of the previous attempts to synthesize meta-analyses, and then introduces some of the major overall findings from the synthesis of the 800+ meta-analyses.

Would it not be wonderful if we could create a single continuum of achievement effects, and locate all possible influences of achievement on this continuum? Figure 2.1 shows one possible depiction of this continuum.

Influences on the left of this continuum are those that decrease achievement, and those on the right increase achievement. Those near the zero point have no influence on achievement outcomes.

The next task was to adopt an appropriate scale so that as many outcomes as possible from thousands of studies are converted to this single scale. This was accomplished using effect sizes, and this scale has been among the marvelous advances in the analysis of research studies over the past century. An effect size provides a common expression of the magnitude of study outcomes for many types of outcome variables, such as school achievement. An effect size of  $d = 1.0$  indicates an increase of one standard deviation on the outcome—in this case the outcome is improving school achievement. A one standard deviation increase is typically associated with advancing children's achievement by two to three years, improving the rate of learning by 50%, or a correlation between some variable (e.g., amount of homework) and achievement of approximately  $r = 0.50$ . When

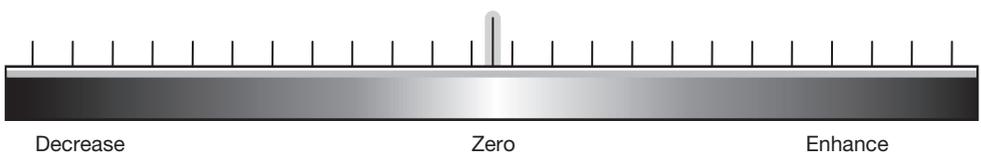


Figure 2.1 An achievement continuum

implementing a new program, an effect size of 1.0 would mean that, on average, students receiving that treatment would exceed 84% of students not receiving that treatment.

Cohen (1988) argued that an effect size of  $d = 1.0$  should be regarded as a large, blatantly obvious, and grossly perceptible difference, and as an example he referred to the difference between the average IQ of PhD graduates and high school students. Another example is the difference between a person at 5'3" (160 cm) and 6'0" (183 cm)—which would be a difference visible to the naked eye. The use of effect sizes highlights the importance of the magnitude of differences, which is contrary to the usual emphasis in much of our research literature on statistical significance. Cohen (1990) has commented that “under the sway of the Fisherian scheme [or dependence on statistical significance], there has been little consciousness of how big things are ... science is inevitably about magnitudes ... and meta-analysis makes a welcome force toward the accumulation of knowledge” (pp. 1309–1310).

Thus, we have a continuum and a scale (effect size) to ascertain which of the many possible influences affect achievement. Many textbooks detail how effect sizes can be calculated from various summary statistics such as  $t$ -tests, ANOVAs, repeated-measures (e.g., Glass, 1977; Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985). Statistically, an effect size can be calculated in two major ways:

$$\text{Effect size} = [\text{Mean}_{\text{treatment}} - \text{Mean}_{\text{control}}]/\text{SD}$$

or

$$\text{Effect size} = [\text{Mean}_{\text{end of treatment}} - \text{Mean}_{\text{beginning of treatment}}]/\text{SD}$$

where SD is the pooled sample standard deviation. There are many minor modifications to these formulas, and for more detail the interested reader is referred to Glass, McGaw, & Smith (1981); Rosenthal (1991); Hedges & Olkin (1985); Hunter & Schmidt (1990); and Lipsey & Wilson (2001).

As an example of synthesizing meta-analyses, take an examination of five meta-analyses on homework: Cooper (1989; 1994); Cooper, Robinson, & Patall (2006); DeBaz (1994); Paschal, Weinstein, & Walberg (1984). Over these five meta-analyses there were 161 studies involving more than 100,000 students, which investigated the effects of homework on students' achievement. The average of all these effect sizes was  $d = 0.29$ , which can be used as the best typical effect size of the influence of homework on achievement. Thus, compared to classes without homework, the use of homework was associated with advancing children's achievement by approximately one year, improving the rate of learning by 15%, about 65% of the effects were positive (that is, improved achievement), 35% of the effects were zero or negative, and the average achievement level of students in classes that prescribed homework exceeded 62% of the achievement levels of the students not prescribed homework. However, an effect size of  $d = 0.29$  would not, according to Cohen (1988), be perceptible to the naked eye, and would be approximately equivalent to the difference between the height of a 5'11" (180 cm) and a 6'0" (182 cm) person.

Thus it is possible to devise a unidimensional continuum such as shown in Figure 2.1 that can allow the various effects on achievement to be positioned as they relate to each other. The scale is expressed in effect sizes (or standard deviation units) such that 1.0 is an unlikely—although a very obvious—change in achievement, and 0.0 is no change at all.

This continuum provides the measurement basis to address the question of the relative effects of many factors on achievement.

An alternative way of considering the meaning of an effect size was suggested by McGraw and Wong (1992). They introduced a measure called the common language effect size indicator, which is the probability that a score sampled from one distribution will be greater than a score sampled from some other distribution. Consider as an example the difference in height of the average woman (5'4"/162.5 cm) and the average male (5'10"/177.5 cm), which is a  $d$  of 2.0. This  $d$  translates into a common language effect (CLE) of 92 percent. Thus we can estimate that in any random pairing the probability of the male being taller than the female is  $d = 0.92$ ; or that in 92 out of 100 blind dates the male will be taller than the female. Now, using the example above, consider the  $d = 0.29$  from introducing homework (throughout this book effect sizes are abbreviated, following tradition, to  $d$ ). The CLE is 21 percent, so that in 21 times out of 100, introducing homework into schools will make a positive difference, or 21 percent of students will gain in achievement compared to those not having homework. Or, if you take two classes, the one using homework will be more effective 21 out of a 100 times. In all examples in this book, the CLE is provided to assist in interpreting the effect size.

We do need to be careful about ascribing adjectives such as small, medium, and large to these effect sizes. Cohen (1988), for example, suggested that  $d = 0.2$  was small,  $d = 0.5$  medium, and  $d = 0.8$  large, whereas the results in this book could suggest  $d = 0.2$  for small,  $d = 0.4$  for medium, and  $d = 0.6$  for large when judging educational outcomes. In many cases this would probably be reasonable, but there are situations where this would be just too simple. Consider, for example, the effects of an influence such as behavioral objectives, which has an overall small effect of  $d = 0.20$  (see Chapter 9), and reciprocal teaching, which has an overall large effect of  $d = 0.74$ . It may be that the cost of implementing behavioral objectives is so small that it is worth using them to gain an influence on achievement, albeit small, whereas it might be too expensive to implement reciprocal teaching to gain the larger effect. Instead of considering only the size of an effect, we should be looking for patterns in the various effect sizes and the causal implications across effect sizes, and making policy decisions on an overall investigation of the differences in effect sizes.

Further, there are many examples that show small effects may be important. A vivid example comes from medicine. Rosenthal and DiMatteo (2001) demonstrated that the effect size of taking low dose aspirin in preventing a heart attack was  $d = 0.07$ , indicating that less than one-eighth of one percent of the variance in heart attacks was accounted for by using aspirin. Although the effect size is small, this translates into the conclusion that 34 out of every 1,000 people would be saved from a heart attack if they used low dose aspirin on a regular basis. This sounds worth it to me.

Meyer *et al.* (2001) list other seemingly small effect sizes with important consequences: the impact of chemotherapy on breast cancer survival ( $d = 0.12$ ), the association between a major league baseball player's batting average and success in obtaining a hit in a particular instance at bat ( $r = 0.06$ ), the value of antihistamines for reducing sneezes and a runny nose ( $d = 0.22$ ), and the link between prominent movie critics' reviews and box office success ( $d = 0.34$ ).

Even more interestingly, it can be possible to identify various moderators that may enhance or detract from the overall average effect. For example, to use the homework case discussed above, it may be that males have greater improvements (i.e., have a higher effect