There have various criticisms of the data in VL, most from the blog space and these also include a lot of not very specific criticism (which are harder to provide a reaction). I think I can claim that I have responded to every person who has emailed me about the book, although I do not respond to personal attacks. Critique is the essence of all academic pursuits and as I noted in VL, I have provided a story that should be subject to falsification and the usual premises of critical investigation. Yes, there have been discussions about the meta-analyses and data but since I produced the first versions of the synthesis (in the late 1980's) no one has come up with an alternative explanation for these data. There have been reviews in peer reviewed Journals, and some scholars have also provided critique in these journals. I welcome such critique.

This brief is to address some of the criticisms of the *methodology*. I welcome any critique and any errors in any of the books – please let me know. This list has been developed from various blog and unpublished comments.

1. The existence of effect-sizes

Some have questioned the existence of effect-sizes and meta-analysis but this methodology is well established. Indeed, the effect-size statistics have become most sophisticated and there are many books on this topic.

Böhning, D., Kuhnert, R., & Rattanasiri, S. (2008). Meta-analysis of binary data using profile likelihood. Boca Raton, FL: Chapman & Hall/CRC.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. Chichester, UK: Wiley.

Cook, T.D., Cooper, H., Cordray, D.S., Hartmann, H., Hedges, L.V., Light, R.J., Louis, T.A. & Mosteller, F. (1992). Meta-analysis for Explanation: A Casebook. New York: Russell Sage Foundation.

Cooper, H. & Hedges, L.V. (Eds.) (1994). The Handbook of Research Synthesis. New York: Russell Sage Foundation.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.)(2009). The handbook of research synthesis and meta-analysis (2nd ed.). New York: Russell Sage Foundation.

Hartung, J., Knapp, G., & Sinha, B. K. (2008). Statistical meta-analysis with applications. Hoboken, NJ: Wiley.

Hedges, L.V. & Olkin, I. (1985). Statistical Methods for Meta-analysis. New York: Academic Press.

Lipsey, M.W. & Wilson, D.B. (2001). Practical Meta-analysis. Thousand Oaks, CA: Sage

Schmidt, F. L. & Hunter, J. E. (2015). Methods of meta-analysis: Correcting error and bias in research synthesis (3rd Ed.). Sage.

Schulze, R. (2004). Meta-analysis: A comparison of approaches. Hogrefe & Huber Pub.

The bases of the method are rather straight forward and indeed much of the usefulness of meta-analysis is its simplicity (see http://www.leeds.ac.uk/educol/documents/00002182.htm). The basic unit is the effect-size – which is akin to standard deviation units, to z-scores and they are the basis of many statistical methods. There are probably not many statistical books that do not mention these measures in some form.

Some have claimed that Mathematicians don't use effect-sizes, Mathematics textbooks don't teach it, and statistical packages don't calculate it. First, effect-sizes are primarily within the discipline of statistics not mathematics; the method is common in statistics text books, and the most widely used statistical packages do calculate it (and there are so many web pages showing how to calculate them)! It is important to note, however, that there are many ways to express an effect-size (eta-squared, and many controversy's such as the optimal estimate of the pooled variance in the

denominator) and many options (use of random or fixed effect models, Bayesian methods etc.). Effect-sizes are pretty basic, standard measures and to deny their existence defies imagination.

2. The tweet that *"half the data in VL are incorrect"*

It is tough to respond to this because I never said this. I was at the CEM conference where I supposedly made this claim, it was recorded so this can be checked, and nowhere was such an absurd claim made. I defy anyone to produce evidence of this claim. It is just not correct. Yes, there are some minor issues (see below), but never is it the case that "half the data" are incorrect. It would be absurd to make such a claim as the data are based on an estimate of effect-sizes from (now) over 1100+ meta-analyses and saying half were incorrect would be accusing these 1100+ authors of error. Let it be clear – I did NOT say this comment.

There are some minor errors in VL, and these are being corrected in the next edition after which they are brought to my attention. I regret these errors, although NONE of them reduce the impact of the story that underlies the VL messages.

a. *The common language effect size estimates in VL are incorrect.*

Yes this is the case in the earlier editions and I regret this. I have issued corrections to everybody who has asked and as new editions and translations come out these have been fixed. It is interesting to note, however, that hardly anyone seems to have used them – it took three years before some students from Norway pointed this out to me and I immediately responded acknowledging this error For those who want, the correct conversions are available but the book can be read by ignoring them – in retrospect they added nothing (even if they had been correct). [Yes, at the last minute in editing I substituted the wrong column of data into the CLE column and did not pick up this error; I regret this omission.] In each subsequent edition the references to CLE and their estimates will be dropped – with no loss to the story.

b. *Inappropriately adding effect-sizes*

Some have claimed that you cannot merely average effect-sizes as this ignores possible moderating influences. There is no doubt that one of the most fascinating aspects of meta-analysis is the opportunity to evaluate the influence of moderators. Indeed the search for moderators has long dominated the educational research literature (e.g., see Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington). For example, does ability grouping work differently in math from music, for 5 year olds compared to 15 year olds? This search, for what is commonly called Aptitude-Treatment Interactions, is as old as our discipline. I actively sought these as finding them would indeed be powerful – as very few have been reported, and hardly any replicated. But the search must continue. As I noted in VL there were very few, and where they did exist I pointed them out (e.g., the differential effects of homework in elementary and high school). If there is no evidence for moderators, then the average across all moderators can be used to make statement about the influence.

Similarly, there is an established methodology about whether the variance of the effects is so heterogeneous that the average may not be a good estimator. Conducting such tests is basic practice in meta-analyses and I implored readers to go to the original studies to see these analyses. I did include an estimator of the variance within each influence (see the dial for each influence) and appropriately commented when these were large. I have spent much time studying many of the

influences with large variance (e.g., feedback) and the story is indeed more nuanced that reflected in the average effect.

c. *Calculating variance with each influence*

There is a rich literature on calculating variance associated with each influence. This is a major concern when conducting a meta-analysis and the methods include evaluating the degree of heterogeneity across the studies, and assessing whether the mean is a reasonable typical measure. The most important consideration is then the search for moderators or mediators to better help explain what is happening across the studies.

Takkouche, B., Khudyakov, P., Costa-Bouzas, J., & Spiegelman, D. (2013). Confidence intervals for heterogeneity measures in meta-analysis. *American journal of epidemiology*, kwt060.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Huedo-Medina, T. B., Sánchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis. *Psychological methods*, *11*(2), 193.

d. *Effect-sizes for gain and comparing two groups.*

There are two major types of effect-sizes – those based on comparing groups (like a class that received a treatment [such as reciprocal teaching] and another class not receiving the treatment). The other is change over time (post-pre). Yes, they have different interpretations. But it is an empirical question as to whether they differ. Most of the effects in VL are based on group comparisons, and the effect of the two methods of comparisons was not meaningful (this is an empirical conclusion). Perhaps in the next edition I should identify which method is used, but having looked closely already this will not influence the overall messages.

e. *The use of .4 as a hinge point.*

The use of .4 is easy – it is the average across all meta-analyses. Simple.

Contrary to some claims, these effects are primarily based on real students in real classrooms, using tests created by teachers or by standardised tests. And yes, teachers can and do get effects greater than .4! Certainly in our work in schools we see this on many occasions and indeed we work in schools where the majority of teachers can get these effects. It is an average, not a tattooed amount that must be exceeded! That this average has not changed from when I first published the studies (in 1989) to today is a test of the robustness of this average.

Of course, care is needed in using this .4. As can be shown, if the tests are measuring a narrow concept then the effects can be higher than if they are measuring a broad concept. The effects can be higher in junior compared to senior classes (often because of the narrow to broader notion). And yes any effect is only as robust as the measures they are based on. Yes, the time over which any intervention is conducted can matter (we find that calculations over less than 10-12 weeks can be unstable, the time is too short to engender change, and you end up doing too much assessment relative to teaching). These are critical moderators of the overall effect-sizes and any use of hinge=.4 should, of course, take these into account.

In many schools we work in the average is much greater than .4, and oftentimes we use this as the hinge-point to then better understand the nature of success and impact of programs.

It is interesting to note that an independent estimate (that is, not from the study of meta-analyses in VL) indicates that the average growth per year in math and reading is about .40. I derived this from an analysis of data from NAEP (USA), NAPLAN (OZ) , SaTS (UK), and e-asTTle (NZ). So we do have a sense of average change – but again it is *average* and as noted above always critical to look for moderators (it is more like .5 in primary and .3 in secondary schools – for these estimates of narrow excellence). It is the interpretation that matters.

f. Should we ditch Low and only use High effects?

I thought I went to great lengths to emphasize interpretations and yes the league table may not have helped (it was meant as an organiser and certainly worked in this respect). I have continued to publish on the influences that intrigue me – particularly why are some effects so low (see below for some references). For example, the effects of class-size are much lower (but positive) than many have argued. I have explored this at length and tried to a) accept the evidence that the effects are small and b) understand why they are small (Hattie, J.A.C. (2007). The paradox of reducing class size and improved learning outcomes. *International Journal of Education, 42*, 387-425). I think we have some good reasons to understand why there are low effects from class size. Similarly, I have embarked on a 10 year search as to why the effects of teacher subject matter are so low. I do not like this conclusion, but can accept the results. Our team is close to understanding this, and it is fascinating to study findings contrary to expectations.

At the other end, I am still working on the influence of feedback – it is among the highest but also most variable effects (lots of feedback is positive and much is negative) – this is an effect where the variance is critical. I have come to understand this variance and still working on this important moderators and influences relating to feedback. For example, it is critical to distinguish between giving and receiving feedback, between "how am I going" and "where the next" feedback, and the feedback effects from students to teachers. More to come on this important topic.

Yes, in some cases small effects can be valuable – they can indicate that the intervention is moving in the right direction, they can indicate that some deeper processes may be changing, and they can indicate that more time, implementation press, or adjustment is needed. Just because an effect is not >.40 does not mean it is not worthwhile. Much more context is needed. Much more interpretation is needed. Statistics help thinking they do not replace thinking. They guide, surprise, correct and suggest – they are not ends in themselves. For example, if I was trying to change conceptual understanding in music theory I might be pleased with d =.2, but maybe not if I was trying to change science vocabulary before starting some lessons in conceptual change. It depends.

g. Other errors

Occasionally people email me with minor errors in VL and I have collected these in anticipation of a second edition of VL (maybe for next year). None are substantial but they will be corrected. If anyone notices any errors I would welcome being told.

3. **The studies in VL**

a. He includes studies of low quality, such as dissertations.

I did note in VL that I was not going to delve deeply into the issue of coding the meta-analysies for quality. This has been misinterpreted as claiming I was not interested in quality -- this is a superficial reading of what I said. I noted that others had conducted this research and given it would have lengthened the book appreciately there seemed little reason to rehearse these claims. It is an empirical question whether the quality of study is a moderator. As I stated in VL: Lipsey and Wilson (1993), for example, summarized 302 meta-analyses in psychology and education, and found no differences between studies that only included random versus non-random design studies (*d* = 0.46 vs. *d* = 0.41), or between high (*d* = 0.40) and low (*d* = 0.37) quality studies. There was a bias upwards from the published (*d* = 0.53) compared to non-published studies (*d* = 0.39), although sample size was unrelated to effect size (*d* = -0.03). Sipe and Curlette (1996) found no relationship between the overall effect size of 97 meta-analyses (*d* = 0.34) and sample size, number of variables coded, type of research design, and a slight increase for published (d = 0.46) versus unpublished (d = 0.36) meta-analyses. There is one exception which can be predicted from the principles of statistical power: if the effect sizes are close to zero, then the probability of having high confidence in this effect is probably related to the sample size (see Cohen, 1988, 1990). ... The aim should be to summarize all possible studies regardless of their design—and then ascertain if quality is a moderator to the final conclusions. I did, where appropriate, note when quality was a moderator and this is forgotten by the critics who misread the claims about quality.

The criticism implied that all dissertations are of lower quality. This is unreasonable – sometimes these studies are better in quality than many published studies (see the lack of difference between published and unpublished studies in the Lipsey and Sipe analyses), and it is possible to use publication as a moderator – and those that have have rarely found this to make a difference.

b. Questionable groupings of meta-analyses.

Yes of course this can be questioned – and I provided all the data such that others could group in other ways, if they wished. The text in VL provides much defense of my grouping. In many cases the details within each grouping is critical. For example, the research on feedback needs much within-group sorting and this has led me and others to many key interpretations.

Perhaps there may be some meta-analyses that need removing. Agreed, and I have removed, for example, the Datta and Marayanan meta , and I have decided to remove quite a few of the gender effects – removing these has not changed the messages or story. These are minor corrections.

c. Combining apples and oranges

This is a common criticism of meta-analyses and been dealt with in many other sources. Simply, combining apples and oranges is common and we call it fruit salad. But the combining remains critical and this is why care is needed to discover moderators and mediators. I searched continually for moderators and there were very few – but where they did exist they were mentioned. Again, the text in VL points this out – especially for feedback where there is considerable variance and as I said much more is needed. Some generic criticisms miss the basis of how to do a meta-analysis.

One concern that is more important is combining two quite divergent effects and assuming the average is a good measure of the "typical value". For example, in "Inductive Teaching" I combine two meta-analyses with effect sizes of d = 0.06 and d = 0.59 to a mean effect size of d = 0.33. But this is a misleading criticism as this was noted in the text – such critics commit their own sin of only looking at the numbers and not the interpretation. I noted that Lott's (1983) meta-analysis included a comparison of inductive versus deductive teaching approaches in science education (where it made little difference) and Klauer and Phye (2008) who were more interested in inductive reasoning

across all subject areas (where they did find higher effects). Of course the details of the interventions matter crucially – which is why the story underlying the data is so critical.

d. The ranking of effects

Yes the rankings of effects have attracted much comment. I took much pain, however, to point to the importance of the underlying story (making learning visible, and since the 2009 book, Know thy impact) and the many nuances around this story. I have taken pains to point to the importance of fixing some of those with low effects (e.g., teacher subject matter knowledge), to understand some lower than expected (e.g., class size), and those with much vairance (e.g., feedback) – and have continued to research and pubish on these topics. The ranking is far less critical than the underlying story.

e. Effect-sizes are not as good as analysisng original data

This is so correct but misses the point. The reason why Glass distinguished between primary, secondary and meta-analysis is that often the original data are not available. If I had original (primary) data I would have used these; if I had access to secondary (access to primary data and conducting re-analyses) I would have used these; but I had no such access – hence the power of meta-analysis. Meta-analysis has allowed me to point to some critical and fascinating questions that need follow up study and (as far as time and resources permit) I have engaged in many further primary studies based on fascinating findings.

There is a whole compendium of risks to interpretation to statistics and these are not unique to meta-analyses. For example, Black and Wiliam (1998a) noted, effect size is influenced by the range of achievement in the population. "An increase of 5 points on a test where the population standard deviation is 10 points would result in an effect size of 0.5 standard deviations. However, the same intervention when administered only to the upper half of the same population, provided that it was equally effective for all students, would result in an effect size of over 0.8 standard deviations, due to the reduced variance of the subsample. An often-observed finding in the literature—that formative assessment interventions are more successful for students with special educational needs (for example in Fuchs & Fuchs, 1986)—is difficult to interpret without some attempt to control for the restriction of range, and may simply be a statistical artefact." But this problem with restriction of range can occur in primary, secondary and meta-analyses. Those of us brought up with Campbell and Stanley (1963, *Experimental and quasi-experimental designs for research on teaching*. American Educational Research Association) will know of many other possible threats to the validity of interpretation of statistics, no matter whether primary or meta. Criticising meta-analyses as if these artefacts are singular to them is misleading.

I would also worry (and cited this in VL) about the narrow or width of the outcome (it is easier to gain higher effects on narrow than wide measures), age (it is more likely that elementary school students gain higher effects on reading and writing than high school students – but care is needed in generalising this to all subject), and surface vs. deep (it is easier to get higher effects on surface than deep measures). Other moderators include nature of students (struggling, medium and high ability), sensitivity to instruction, country (developing and developed) and so on – this is basic social science caution and all noted in VL.

f. Provide clarity around the process used to ensure that assumption of normal distribution was achieved, and the same/similar standard deviations were found across studies – and were studies only averaged where the measurements used had a known/reported reliability.

There is no requirement in calculating and using effect-sizes as I have done in VL that requires the assumption of normal distribution. If (and I do not) make statistical probability statements maybe, but this is not how effect-sizes are used. Similarly there is no requirement that sd is the same across studies – they depend more on the scale of the measures used within studies. As the requirement of measurement reliability (and I presume validity) many meta-analyses coded and analysed their studies for "quality" (and often these included the quality of the measures). Where there were no difference in quality all were included, otherwise the low quality studies were more often deleted.

g. *Please explain the reason why confidence intervals were not used to help convey the effect size information.*

Yes, they could be used and most users of meta-analyses would know how to calculate these confidence intervals from the information I provided in VL. I decided it was an unnecessary extra as it implied a false sense of precision. The average effects needed careful interpretation and where appropriate commentary about the moderators that seemed to affect the average.

h. *Explain the methodology around the establishment of the .4 average across all of the meta-analyses and the justification for averaging effects where standardised and unstandardized (e.g., teacher's own) tests have been used.*

Simple, the .4 is the average across the 800 meta-analyses. The justification is that, where there was information within a meta-analysis that investigated the differences between standardised and unstandardized (e.g., teacher's own) tests there were no differences. Much more important, as I noted in VL, was whether the measure was for a narrow or a broad dimension (e.g., phonics vs. comprehension) and where these occurred I made appropriate comments.

i. *Provide references to the published studies you refer to, that span across 1989 to now that reflects empirically the justification for the .4 average.*

All the meta-analyses I used in VL are referenced in the books. None of these, of course justify the .4 average as this is what I found by averaging as noted above. As I note in many places, an average is just that- -- and care is needed to dig below the numbers. The .4 serves as a hinge point such that I then tried to explain the "story" as to what was most related to the effects above compared to below this average. The quality of the story is the key contribution.

j. *Provide clarity around how the league table of effects sizes should be interpreted, to assist in understanding that influences that fall below .4 are still worthwhile.*

This is the primary purpose of Visible Learning, and Visible Learning for teachers. It will also be explained in the next book (due April 2015) where we look at 15 case studies of schools that have implemented the story underlining VL – and you will see that there is no one way, no recipe, no script, no workbook – but a constant attention to evaluating the impact of all in a school. This begs the question as to whether all in the school share a common notion of "impact" (and if impact is simply test scores you fail in my opinion; and if it does not you are missing the obvious), an understanding of how to estimate the magnitude of the impact (and if your only use effect-sizes you are too narrow), and the pervasiveness of impact (across how many students).

Conclusions

I welcome any and every correction, missing meta-analyses, and critique of my work. Yes, the academic world of critique is changing as the blog space is more present. Often this makes reactions difficult, especially when the same blogs enter in personal attacks. It does not matter my background (I am a psychometrician with a Phd in measurement; past-President of the International Testing Commission and so on), what matters is the corpus of what I say and the defence of these claims. Sometimes the personal attacks bemuse but mostly they are best ignored. For example I have a BA, and a PhD (in educational theory from the University of Toronto). Yes, my BA includes statistics and my Phd was in Psychometrics. But the vagaries of how Universities label degrees can be interesting. I make no claim for expertise in the Arts (although I was once a music teacher) or educational theory (although fascinated with the philosophy of science) so to be accused of some Arts graduate with no knowledge of basic statistics is more amusing than facutal.

I did attend lectures while a student at the Unviersity of Otago by Sir Karl Popper – and was most influenced (and still am) by his theories based on falsification. Right now, I stand by the story developed in Visible Learning until some one comes up with an alternative explanation that better explains the evidence and falsifymy interpretations. I continue to research, modify (especially in light of our teams work in schools) the messages, and welcome critique (oh, to be ignored!). Yes the details are important and this is why I continue to explore these – for example every influence is discussed in some detail in VL; and more recently Eric Anderman and I edited a book asking over 200 academcis who specifically research on the many influences to write a chapter in this book (Hattie, J., & Anderman, E. M. (Eds.). (2012). *International guide to student achievement*. Routledge). We have explored the meaning of the story in our work on teaching Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge), learning (Hattie, J., & Yates, G. C. (2013). *Visible learning and the science of how we learn*. Routledge), and soon a book on *VL in action* in schools (Hattie Masters & Birch, Routledge, 2015). I have published many more detailed studies on many of the influences in peer reviewed journals and invite others to question, test, and develop interpretations about the overall messages. The quest continues.

None of the critiques above dent the major stories. The main message remains, be cautious, interpret in light of the evidence, search for moderators, take care in developing stories, welcome critique, and critique the ideas not the person.

I have continued to publish on many aspect of the work in VL .. and now working on a synthesis of studies relating to learning strategies, advising a team based in Germany on a synthesis relating to motivation and affective outcomes, applaud David Mitchell's book synthesising the research relating to special education students (http://www.routledge.com/cw/mitchell-9780415623230/). Some publications over the past few years include:

Books and monographs:

1. Hattie, J.A.C. & Yates, G. (2014). *Visible Learning and the Science of how we Learn*. Routledge, UK.
    a. Norwegian version: *Synligt larande: hvordan vilaerer.* Cappelen Damm, 2014.
    b. Danish version: *Synlig laering og laeringens anatomi*. Dafolo. 2014.
2. Hattie, J.A.C., & Anderman, E. (2013). *Handbook on Student Achievement*. Routledge, New York.
3. Hattie, J.A.C. (2012). *Visible learning for teachers*. *Maximizing impact on achievement*. Oxford, UK: Routledge.

a. Swedish version: *Synligt larande for larare.* Natur & Kultur, 2013.

b. Danish version: *Synligt larande for larare.* Dafolo, 2013.

c. Norwegian version: *Synligt larande for larare.* Dafolo, 2013.

d. Dutch version: *Leren zichtbaar maken*. Bazalt, 2013.

e. German version: *Lernen sichtbar machen fur Lehrpersonen*. Schneider Verlag Hohengehren GmbH. Trans Wolfgang Beywl & Klaus Zierer

4. Hattie, J.A.C. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement.* Oxford, UK: Routledge.

    a. German version: *Lernen sichtbar machen*. Wolfgang Beywl & Klaus Zierer (Trans). Schneider Verlag Hohengehren GmbH, 2013.

    b. Swedish version: Synligt larande. Natur & Kultur, 2013.

    c. Swedish version: Synligt larande. Natur & Kultur, 2014. Second edition

Papers

5. Hattie, J.A. C. (in press). High-Impact Leadership: Effective instructional leaders don't just focus on student learning. They relentlessly search out and interrogate evidence of that learning. *Educational Leadership.*

6. Pfost, M., Hattie, J.A.C., Dorfler, T., & Artelt, C. (2014). Individual Differences in Reading Development: A Review of 25 Years of Empirical Research on Matthew Effects in Reading, *Review of Educational Research, 84(2*), 203-244.

7. Gan, M.J.S., & Hattie, J.A.C. (in press). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*.

8. Clinton, J.M., & Hattie, J.A.C. (2014). Education and Empowerment Evaluation. In Fetterman, D., Kaftarian, S., and Wandersman, A. (Eds.). *Empowerment Evaluation: Knowledge and Tools for Self-assessment, Evaluation Capacity Building, and Accountability*. Thousand Oaks, CA: Sage.

9. Hattie , J.A.C., & Wollenschläger, M. (2014). *A conceptualization of feedback*. In H. Ditton & A. Muller (Eds)., Feedback und Ruckmeldungen. (pp. 135-150) Munster Germany: Waxmann.

10. Wollenschläger, M., Hattie, J.A.C., Möller, J., & Harms, U. (in review). Competential Feedback Effects on Performance: Are They Mediated by Calibration?

11. Hattie, J.A.C. & Yates, G.C.R. (2014). Using feedback to promote learning. In V.A. Benassi, C.E. Overson, & C.M Hakala (Eds). *Applying science of learning in education: Infusing psychological science into the curriculum*. pp. 45-58. Washington, DC, American Psychological Association. http://teachpsych.org/ebooks/asle2014/index.php

12. Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2013). The effects of feedback on achievement, interest, and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology, 24(4),* 269-290. DOI**:** 10.1080/01443410.2013.785384

13. Clinton, J.C., & Hattie, J.A.C. (2013). New Zealand students' perceptions of parental involvement: Relations with liking, efficacy and achievement. *Asia Pacific Journal of Education, 33(3)*, 324-337.

14. Yates, G.C.R., & Hattie, J.A.C. (2013). Experts amongst us: What do we know about them? *Journal of Educational Enquiry*, 12(1), 40-51.

15. Suri, H., & Hattie, J.A.C. (2013). Meta-analysis and research synthesis in Education. *Oxford Bibliographies*. http://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0091.xml?rskey=U1CGQW&result=1&q=hattie%27#firstMatch

16. Winheller, S., Hattie, J.A., & Brown, G. (2013). Factors influencing early adolescents' mathematics achievement: High-quality teaching rather than relationships. *Learning Environments Research, 16(1),* 49-69. doi: 10.1007/s10984-012-9106-6

17. Suri, H., & Hattie, J (2013). Meta-Analysis and Research Synthesis in Education. In L. Meyer (Ed.), Oxford *Bibliographies in Education*. New York: Oxford University Press.

18. Hattie, J.A.C. (2013). Calibration and confidence: Where to next? *Learning and Instruction, 24*, 62-66.

19. Hattie, J.A.C . (2013). The power of feedback in school settings. In R. Sutton (Ed). *Feedback: The handbook of criticism, praise, and advice.* Peter Lang.

20. Hattie, J., & Clinton, J. (2012). Physical activity is not related to performance at school. *Archives: Pediatrics and Adolescent Medicine, 166(7*), 678.

21. Watkins, D.W. & Hattie, J.A.C. (2012). Multiple Goals in a Hong Kong Chinese Educational Context: An Investigation of Developmental Trends and Learning Outcome, *Australian Journal of Education, 56(3*), 273-286.

22. Hattie, J.A.C. (2012). Know thy impact. *Educational leadership, 70(1),* 18-23.

23. Chen, J.J., Brown, G.T., Hattie, J.A.C., & Milward, P. (2012). Teachers' conceptions of excellent teaching and its relationship to self-reported teaching practices. *Teaching and Teacher Education*, *28(7*), 936-947. doi:10.1016/j.tate.2012.04.006

24. Bendiksen, L., Robinson, V.M.J., & Hattie J.A.C. (2012). Principal Instructional Leadership – what is it and what is its relationship to secondary school performance? *SET: research information for teachers, 1*, 2-8.

25. Hattie, J.A.C., & Gan. M. (2011). Instruction based on feedback. In R. Mayer & P. Alexander (Eds.), *Handbook of Research on Learning and Instruction*. (pp. 249-271). New York: Routledge.

26. Hattie, J.A.C. (2011). Purposes and uses of educational research. *QTU Professional Magazine, 26*, Nov, 8-10.

27. Hattie, J.A.C., & Clinton, J.M. (2011). School leaders as evaluators. In *Activate: A leader's guide to people, practices and processes*. (pp. 93-118). Englewood, CO: The Leadership and learning Center.

28. Robinson, V.M.J., Bendikson, L., & Hattie, J.A.C. (2011). Leadership and student outcomes: Are secondary schools unique? In J. Robertson & H. Timperley (Eds.), *Leadership and learning.* (pp. 131-142). Los Angeles: Sage.

29. Hattie, J.A.C., Rogers, H.J., & Swaminathan, H. (2011). The role of meta-analysis in educational research. In M. Peters (Ed.), *Contextualising Educational Research*. Sage: Beverly Hills, CA.

30. Bendiksen, L., Robinson, V.M.J., & Hattie, J.A. (2011). Identifying the comparative academic performance of secondary schools. *Journal of Educational Administration*, *49,* 433-449.

31. Mhurchu, C.N., Turley, M., Gorton, D., Jiang, Y., Michie, J., Maddison, R., & Hattie, J. (2011). Effects of a free school breakfast programme on school attendance, achievement, psychosocial function and nutrition: A stepped wedge cluster randomised trial. *BMC Public Health,* 10. http://www.biomedcentral.com/1471-2458/10/738

32. Hattie, J.A.C. (2011). Which Strategies Best Enhance Teaching and Learning in Higher Education? In D. Mashek & E. Hammer (Ed.), *Empirical research in teaching and learning: Contributions from social psychology.* (pp. 130-142). Wiley-Blackwell.

33. Hattie, J.A.C. (2010). On being a 'critic and conscience of society': The role of the education academic in public debates. *New Zealand Journal of Educational Studies, 45(1),* 85-96.

34. Hattie, J.A.C. (2009). The black box of tertiary assessment: An impending revolution. In L.H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P.M. Johnston, & M. Rees, *Tertiary assessment and higher education student outcomes: Policy, practice, and research.* Pp. 259-276. Wellington, New Zealand: Ako Aotearoa.

35. Absolum, M., Flockton, L., Hattie, J.A.C., Hipkins, R., Reid, I (2009). *Directions for Assessment in New Zealand: Developing students' assessment capabilities.* Ministry of Education, Wellington, NZ. http://assessment.tki.org.nz/Assessment-in-the-classroom/Directions-for-assessment-in-New-Zealand-DANZ-report

36. Hattie, J.A.C. (2009). Narrow the gap, fix the tail, or close the curves: The power of words. In Rubie, C., & Rawlinson, C. (Eds.), *Challenging Thinking about Teaching and Learning.* Nova Science.

37. Carroll, A., Houghton, S., Wood, R., Unsworth, K., Hattie, J., Gordon, L., & Bower, J. (2008). Self-efficacy and academic achievement in Australian high school students: The mediating effects of academic aspirations and delinquency. *Journal of Adolescence, 32,* 797-817. http://www.sciencedirect. com/science?_ob= ArticleURL&_udi=B6WH04V053FV1&_user=10&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=18198a98853b2990f75c7d0e1c641203)

38. Hattie, J.A.C. (2008). The rope model of self. In H.W. Marsh, R.G. Craven, & D.M. McInerney (Eds) *Self-Processes, Learning, and Enabling Human Potential: Dynamic New Approaches* (pp.56-86). Charlotte, NC: Information Age Publishing, Inc.

39. Hattie, J.A., & Clinton, J.C. (2008). Identifying accomplished teachers: a validation study. In L. Ingvarson & J.A.C. Hattie (Eds). *Assessing teachers for professional certification: The first decade of the National Board for Professional Teaching Standards*. Pp. 313-344. Advances in Program Evaluation Series #11, Oxford, Elsevier.

40. Hattie, J.A.C., & Timperley, H (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. Republished *in Feedback log vurdering for laering*. Dafolo.

41. Hattie, J.A.C. (2007). The paradox of reducing class size and improved learning outcomes. *International Journal of Education, 42*, 387-425.

42. Rubie-Davis, C., Hattie, J.A.C., & Hamilton, R.J. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *British Journal of Educational Psychology, 76,* 429-444.

43. Lee, E.W.M., Watkins, D., & Hattie, J.A. (2006). Late change in the medium of instruction: A Hong Kong investigation. *The Asia-Pacific Education Researcher, 15(1), 3-15.* Rreissued in *Educational Reform and Innovation in the Asia-Pacific: Advances in Research* (2013) -- edited by Allan B. I. Bernardo and Marianne Jennifer M. Gaerlan, and published by the De La Salle University Publishing House.

44. Anderson, A., Hattie, J.A.C., & Hamilton, R.J. (2005). Locus of control, self-efficacy and motivation in different schools: Is moderation the key to success? *Educational Psychologist*, 25(5) 517-535.