

Reacties op het essay van Dinand Webbink

Nr. 20130216/1027

Onderwijsraad
Nassaulaan 6
2514 JS Den Haag
e-mail: secretariaat@onderwijsraad.nl
070 – 310 00 000 of via de website: www.onderwijsraad.nl

Inhoudsopgave

1	Reactie Jules Peschar: over doelen en doelmatigheid in het onderwijs	5
2	Reactie Hans Boutellier: Dynamische overdracht	9
3	Reactie Jaap Scheerens	12
4	Reactie Gert Biesta: Educational Efficiency, Educational Aims, and the Complexity of Educational Systems	14
5	Reactie John Hattie	19

1 Reactie Jules Peschar: over doelen en doelmatigheid in het onderwijs

Jules Peschar is emeritus-hoogleraar onderwijssociologie aan de Rijksuniversiteit Groningen

In deze bijdrage wordt vanuit een sociologisch gezichtspunt gereageerd op het interessante en genuanceerde essay van Dinand Webbink *Efficiency in Education*. Mijn reactie bestaat uit enkele algemene observaties, een korte beschouwing over de implicaties, enkele lessen voor het beleid en een antwoord op de door de Onderwijsraad gestelde vragen.

Algemene opmerkingen

Iedere discipline kiest zijn eigen perspectief. Wanneer economen naar het onderwijs kijken, benoemen ze de processen in termen van *productiefuncties*. Sociologen kijken vooral naar de centrale functies van onderwijs. Veelal worden drie functies onderscheiden: in het onderwijs worden leerlingen *gekwalificeerd* (verwerven ze competenties); er wordt *gedifferentieerd* (niet iedereen behoeft dezelfde kwalificaties) en er worden leerlingen *gesocialiseerd* (er vindt overdracht van waarden en normen plaats) (Peschar & Wesseling, 1995).

Wanneer economen over efficiency en doelmatigheid spreken, dan gaat het vrijwel uitsluitend over de kwalificatiefunctie en deze wordt meestal omschreven in termen van behaalde cijfers of scores op een gestandaardiseerde test. Dat is niet alleen een smalle operationalisatie van het kwalificatiebegrip –daarover verderop nog meer—maar tevens is er hiermee weinig oog voor de twee andere functies van het onderwijs. Men zou zich immers ook kunnen afvragen of er effectief wordt gedifferentieerd in het onderwijs en hoe doelmatig de aan het onderwijs toegeschreven socialisatietaak wordt vervuld. Gelukkig heeft Webbink hier in zijn essay wel oog voor, wanneer hij het over non-cognitieve doelen van het onderwijs heeft. Misschien is het goed om deze expliciet te benoemen, want gaat het niet concreet om zaken als sociale competenties, motivatie of burgerschap?

De vraag naar doelmatigheid –of dat nu *input of output efficiency* is – impliceert dat er een doel is gekozen dat op meerdere manieren kan worden bereikt: zonder dat doel valt iets als doelmatigheid niet vast te stellen. Maar ligt juist daar in de huidige onderwijsconstellatie niet het probleem? Hier vindt een opvallende verschuiving in de doelen plaats: er wordt vooral gekozen voor *meetbare* doelen en deze zijn vaak weer gedefinieerd door de meetinstrumenten die beschikbaar zijn: IQ, prestaties of toetsen. Dit beperkt niet alleen het zicht op bredere functie van het onderwijs, maar leidt ook tot strategisch gedrag van scholen en leerkrachten. In een afrekencultuur is het niet onbegrijpelijk dat scholen zich bepalen tot meetbare prestaties en daar het onderwijs op inrichten (teaching-to-the-test).

Een belangrijke veronderstelling bij de vaststelling van doelmatigheid is dat er één of meerdere criteria beschikbaar zijn, waardoor een effect is vast te stellen. Veruit het meeste onderzoek gebruikt een gestandaardiseerde test voor leerprestaties, in de veronderstelling dat de test ook *in de tijd* stabiel is. Vaak blijft buiten beeld dat juist die stabiliteit in de tijd een probleem is. Geregeld worden testen opnieuw gevalideerd en worden nieuwe populatiegemiddelden en standaarddeviaties vastgesteld. In feite zijn dergelijke tests minder geschikt voor onderzoek naar doelmatigheid, waar per definitie een tijdsdimensie in zit. Dit geldt niet alleen voor de Nederlandse peilingsonderzoeken (PPON), maar ook voor de Cito-scores en de toetsen uit de internationale PISA-vergelijkingen.¹ Empirisch onderzoek is noodzakelijk, maar men moet ook oppassen schijnzekerheden te presenteren.

¹ In elke driejaarlijkse ronde van PISA wordt het gemiddelde van de toetsen voor de OESO-landen opnieuw op 500 gesteld (zie onder meer Organisation for Economic Co-operation and Development, 2006, p.369 ff).

Over implicaties

In grote lijnen onderschrijf ik de hoofdargumenten en de implicaties die Webbink aangeeft. Belangrijk is zijn constatering dat over veel beleid niet kan worden geoordeeld; bijvoorbeeld de relatie tussen onderwijsbudget en prestaties. Er blijkt onvoldoende variatie in de relevante variabelen: er is nooit een vermindering van onderwijsbudgetten geweest en dus kan een causaal effect op prestaties niet worden aangetoond (zie zijn figuur 2). In veel gevallen is er onvoldoende onderzoek geweest om een uitspraak te kunnen doen.

De onderstaande kanttekeningen zijn vooral bedoeld ter aanvulling op zijn analyse van pagina 13 en 14.

Hierboven werd ingegaan op de versmalling van onderwijsdoelen tot meetbare kwalificaties. Maar wellicht nog belangrijker is de vraag welke doelen eigenlijk worden nagestreefd. Hoe kun je vaststellen of een doel bereikt wordt, wanneer die doelen vaag, algemeen of zelfs helemaal niet worden geformuleerd?

Neem bijvoorbeeld passend onderwijs. De doelstellingen van het beleid hebben vooral betrekking op de structuur en financiering van onderwijs voor een grote groep – geschat wordt ongeveer een kwart van alle – leerlingen. Waar voor het overige onderwijs expliciete eisen of een kwaliteitskaart van competenties is vastgelegd, worden voor passend onderwijs geen *inhoudelijke* onderwijsdoelen voor leerlingen geformuleerd. De evaluatie van een dergelijk proces – of het op macro-, meso- en microniveau werkt – is dan ook een lastige klus². Vanzelfsprekend wordt een monitor ontwikkeld en worden deelonderzoeken in scholen en samenwerkingsverbanden opgezet. Maar de vraag in hoeverre leerlingen na de invoering van passend onderwijs “beter worden bediend” door de nieuw ingevoerde structuur en middelen kan niet worden beantwoord. Even los van de doelspecificatie zijn er immers geen (semi-) experimenten geweest waarin verschillende varianten werden beproefd (wat goed werkt, dat voeren we in; zo niet, dan doen we het anders). Paradoxaal genoeg waren die experimenten er in beperkte mate aanvankelijk wel, maar werden ze voortijdig door staatssecretaris Dijkma afgeschaft. De richting is al gekozen en er is geen weg meer terug.

Met Webbink kan men het alleen maar eens zijn dat er meer moet worden ingezet op ‘what works’. Dat is in lijn met de aangehaalde meta-analyses van economen en onderwijskundigen. Het probleem hierbij is alleen dat die resultaten niet zo eenduidig zijn als men zou wensen. Men lijkt daarbij overigens – en niet alleen in onderwijsland – vooral gericht te zijn op successen en ‘good practices’. Minstens zo belangrijk is echter de vraag om vast te leggen ‘what works NOT’. Dergelijke resultaten zijn onnoemelijk veel belangrijker voor de praktijk; er wordt mee voorkomen dat het wiel opnieuw wordt uitgevonden en dat reeds verkende doodlopende paden wederom worden onderzocht. Helaas is de documentatie van negatieve bevindingen en ‘bad practices’ niet populair en in wetenschappelijke tijdschriften zijn ze vrijwel niet terug te vinden.

Een belangrijk argument van Webbink lijkt te zijn dat de causale effecten van interventies en innovaties in zo sterk mogelijke designs moeten worden vastgesteld. Daar ben ik het volmondig mee eens. In *abstracto*, want de beleidspraktijk heeft zich hier nogal tegen verzet. Ook politiek liep men niet echt warm voor harde evaluaties en experimenten. De middenschool-experimenten waren weliswaar geen echte experimenten, maar mochten ook niet op andere wijze worden geëvalueerd.³ In grote lijnen geldt dit eveneens voor latere evaluaties van basisvorming, weer samen naar school, en passend onderwijs. Er werden uitvoerige evaluatieprogramma’s opgezet, maar er werden geen mogelijkheden toegestaan om vergelijkingen te maken met groepen die als controlegroep kunnen worden beschouwd. Voorstellen in die richting werden systematisch afgewezen.⁴

² Zie het evaluatieplan van de evaluatiecommissie ECPO, www.ecpo.nl.

³ Dat we hier toch nog iets van hebben kunnen leren is puur toeval. GION-onderzoeker Annemieke de Vries had voor de middenschoolen veel onderzoek gedaan en kreeg het vertrouwen van *die scholen* om de gegevens te vergelijken met een landelijk onderwijscohort, rekening houdend met allerlei relevante doorkruisende variabelen. Ze vond een positief effect van groepeeringsvormen op leerprestaties voor achterblijvende leerlingen. Bij haar promotie vroeg de toenmalige secretaris-generaal van OCW dr. Ferdinand Mertens zich af waarom dergelijk onderzoek niet eerder was gedaan, aangezien het ministerie “hierop had zitten wachten”. Maar datzelfde ministerie had eerder elke harde evaluatie van de Middenschool effectief afgehouden (De Vries, 1992).

⁴ Ik was in de rol van coördinator, voorzitter of lid van de commissie betrokken bij de evaluatie van de drie genoemde onderwijsvernieuwingen.

De recente middenstroom voor Onderwijsbewijs mocht overigens uitsluitend worden gebruikt om gerandomiseerde experimenten in het onderwijs op te zetten en te evalueren. Verschillende pogingen om ook andere harde evaluatiedesigns – met gematchte groepen, veldexperimenten, regressiemodellen (nota bene een stokpaardje van het CPB) of tijdreeksen – toe te staan liepen op niets uit. Met als gevolg dat ook de ‘next-best’ harde evaluaties niet konden worden uitgevoerd, wanneer een experiment onmogelijk bleek.

Lessen voor beleid

Een eerste les zou zijn dat er voor alle beleidsprogramma’s wordt vastgesteld welke concrete doelen worden nagestreefd en voor welke groepen en organisaties dat zou moeten gelden. Te vaak wordt dat in het vage gehouden, waardoor er een misplaatst gevoel van algemene overeenstemming ontstaat.

Als je een algemene regel zou moeten opstellen – een tweede les – dan zou het deze zijn: Als je beleid formuleert, dan moet je *van te voren* vastleggen of en hoe je kunt vaststellen of het doel wordt bereikt. In feite moet de causale keten worden uitgeschreven waarlangs nieuw beleid (een maatregel of vernieuwing) een effect kan hebben op het voorgenomen doel. Pas als er meerdere wegen tot eenzelfde doel zouden leiden, kan iets over doelmatigheid worden gezegd. Ex-ante evaluaties – zoals voor passend onderwijs ondernomen – zouden verplicht moeten zijn (Ledoux, 2012).

Een derde les heeft betrekking op het feit dat veel beleid niet door het ministerie van OCW zelf wordt uitgevoerd, maar in de praktijk wordt gedelegeerd aan schoolbesturen. Daarmee wordt het lastiger om de werkzaamheid van de causale keten te controleren. Niettemin kan een overheid explicieter kwaliteitseisen stellen aan middelen en programma’s die tot een bepaald doel zouden moeten leiden. Een goed voorbeeld is het recente advies van de Onderwijsraad over burgerschap. Aanbeveling 2 betreft het versterken van de kennisbasis van effectief gebleken programma’s. In het verlengde daarvan zou men kunnen denken aan het certificeren van scholen die met dergelijke programma’s werken. Kwaliteit moet blijken (uit onderliggend onderzoek) en het moet lonen (door extra faciliteiten of inspectieoordelen).

De vragen van de Onderwijsraad

Deel je de visie op ‘doelmatigheid’ en ‘doelmatig onderwijs’ die spreekt uit het essay? Welke kansen en/of risico’s zie je?

Antwoord: Het grootste probleem is dat vaak onduidelijk is welke doeleinden worden nagestreefd; doelmatigheid kan dus niet kan worden vastgesteld.

Indien je een andere visie op doelmatigheid in het kader van onderwijs hebt, kun je deze omschrijven?

Antwoord: het gaat m.i. niet om een andere visie, maar op de randvoorwaarden die aan onderwijsbeleid of specifieke maatregelen worden gesteld.

Welke plaats zou een concept als doelmatigheid (zoals door jou voorgestaan) kunnen/moeten hebben in het onderwijsbeleid van de overheid en in het beleid van scholen?

Antwoord: Zoals hierboven uiteengezet, zou ik nationaal willen kunnen (laten) vaststellen of nieuw beleid inderdaad effectiever is dan bestaande praktijk. Dit kan door een zo hard mogelijke evaluatie; pas daarna zou tot invoering kunnen worden besloten.⁵ Voor scholen geldt dat het gebruik van effectief gebleken methoden en praktijken moet worden beloond en – omgekeerd – het niet-gebruiken tot een comparatief nadeel zou moeten leiden.

Deel je de implicaties voor onderwijsbeleid zoals die in het essay zijn omschreven?

Antwoord: ja, maar ik zou zelfs verder willen gaan, zoals boven geschetst.

Literatuur

De Vries, A.M. (1992). *Hoe breder hoe beter?* Groningen: RION.

⁵ Was dit niet ooit een van de belangrijkste conclusies van het rapport *Tijd voor Onderwijs*, de parlementaire commissie onder voorzitterschap van Dijsselbloem (2008)?

- Ledoux, G. (2012). *Ex-ante evaluatie passend onderwijs. Studie in opdracht van de ECPO*. Amsterdam, Kohnstamm Instituut.
- Organisation of Economic Co-operation and Development (2006). *PISA 2006, Science Competencies for Tomorrow's World. Volume 1*. Parijs: OECD.
- Peschar, J.L. & Wesselingh, A.A. (1995). *Onderwijssociologie*. Groningen, Wolters-Noordhoff.

2 Reactie Hans Boutellier: Dynamische overdracht

Hans Boutellier is algemeen directeur van het Verwey-Jonker Instituut en bijzonder hoogleraar veiligheid & burgerschap aan de VU Amsterdam

Dat de doelmatigheid van het onderwijs omhoog moet, lijkt een uitgemaakte zaak. We zijn een kenniseconomie, die het van een hoogopgeleide bevolking moeten hebben. Diploma's beloven bovendien voor individuele kinderen een goede toekomst. En omdat onderwijs een dure investering is, moeten we het maximaal laten renderen. Opbrengstgericht onderwijs! Maar wat is dat precies, en hoe weten we wanneer daar sprake van is? Kun je dat ergens aan zien en kan je het ook meten? Dinand Webbink biedt een voortreffelijk overzicht van de onderwijs-economische *state of the art*. Valt daar nog iets aan toe te voegen?

State of the art

Webbink constateert dat toename in middelen niet per se leidt tot een hoger rendement in het onderwijs. Dat geldt overigens voor veel meer publieke sectoren (Kuhry & De Kam, 2012). Klaarblijkelijk zit er een ingebouwde weerstand in de verbeteringsmogelijkheden. Het probleem is wel dat we niet weten wat er gebeurd zou zijn zonder extra middelen. De wereld is ingewikkelder geworden, de scholen zijn vernetwerkt, de klassen werden diverser en de leerlingen mondiger. Dikke kans dat het onderwijs slechter geworden was zonder extra middelen.

De onderwijsproductiefunctie – de relatie tussen input en outcome – lijkt dus 'overall' niet te zijn verbeterd (Hanushek in Dinand, 2012). Maar vergelijken is lastig. Wie de schoolsituatie van vandaag vergelijkt met die van veertig jaar geleden, vindt enorme verschillen. In feite levert elke vergelijking problemen op. Dit geldt voor de verschillende aggregatieniveaus waarop vergeleken kan worden: land, school of beleid c.q. interventie. Er lijkt sprake te zijn van een vrij fundamentele onvoorspelbaarheid vanwege de enorme hoeveelheid variabelen die in het geding zijn.

Toch is er veel evidentie dat investeringen in 'human capital' de welvaart van zowel het individu als de samenleving ten goede komen. Bezuinigen zou veel kwaad kunnen doen. Het komt er dus op aan de beschikbare middelen zo verstandig mogelijk te verdelen en geen schade aan te richten. Webbink meent dat – gegeven de grote onzekerheid over wat werkt – twee beleidslijnen het meest heilzaam zijn in het optimaliseren van het rendement van investeringen in het onderwijs:

- *Rekenschap*: ontwerp beleid dat voor scholen duidelijke doelen stelt, waarbij tegelijkertijd strategisch gedrag van scholen (uitsluiten van leerlingen en opleiden tot toetsresultaten) wordt tegengegaan. Hij ziet hier een rol weggelegd voor de inspectie.
- *Leerstrategie*: introduceer beleid op een experimentele manier en vergelijkenderwijze, waardoor meer *zicht* kan worden gekregen op de werkzame elementen en een leerproces op gang kan komen.

Kernfunctie en praxis

Webbink's aanbevelingen lijken rationeel, maar zeker weten doen we het niet. De onderwijspraktijk is de westerse wereld sterk verwetenschappelijkt. Daarbij is het medische model behoorlijk dominant geworden. Via experimenteel onderzoek wordt de werkzaamheid van interventies bepaald en op grond van positief resultaat ook in andere situaties geïmplementeerd. Ik ben van mening dat de vraag naar 'efficiency in education' veelal te krampachtig wordt benaderd. Of interventies heilzaam zijn (om eens een ander woord dan doelmatig te gebruiken) is in de werkelijkheid van alledag vaak moeilijk te bepalen:

- experimenteel onderzoek is vaak niet mogelijk;
- professionals hebben een sterke op zichzelf staande invloed;
- voor het beleid is wat werkt slechts één argument tussen andere;

- het gezag van de wetenschap is afgebrokkeld;

Wat werkt is een variant van een grotere vraag: waar doen we goed aan? In de huidige onoverzichtelijke wereld is er veel behoefte aan richtinggevende informatie, maar deze kan niet zonder meer afgeleid worden uit de waarheidsaanspraak van de wetenschap. Beleid, uitvoering en zelfwerkzaamheid zijn steeds meer een gezamenlijke onderneming van betrokkenen, waarbij onderzoekers kunnen proberen een gerationaliseerde onderbouwing te geven van beleidsbeslissingen en methodieken. Dat betekent dat elke uitspraak over wat werkt staat of valt met de doelstelling van de onderneming (aanbeveling 1 van Webbink).

De praktijk leert dat bestuurders en professionals vaak grote moeite hebben met het vaststellen van doelstellingen. Dat is niets om cynisch over te doen. Men weet gewoonweg vaak niet wat men moet willen en hoopt dat de wetenschap uitkomst kan bieden. Daarmee wordt de rol van onderzoek overschat. Wetenschappers komen vaak tot verschillende resultaten (en dat hoort ook zo) en redeneren noodzakelijkerwijze vanuit assumpties. Het voorspellend vermogen in de sociale werkelijkheid is uiteindelijk beperkt. Onderzoek kan niet beslissen over de doelstellingen van beleid en uitvoering. Het kan daarin hooguit faciliterend zijn.

De vraag naar een optimale onderwijsproductiefunctie begint bij de vaststelling van de doelstellingen, wordt gevolgd door een probleem- of kansanalyse en daarna doelgerichte aanpak. Pas dan kunnen uitspraken worden gedaan over doelmatigheid. In de afgelopen decennia is het onderwijsveld – net als veel andere instituties – in verwarring geraakt over zijn specifieke rol. Het komt erop aan het daarover eens te worden. In een complexe wereld komt het erop aan te redeneren vanuit de kernfunctie van de institutie of de organisatie (voor een uitgebreide argumentatie: Boutellier, 2011). Met alle ingewikkeldheid en in alle varianten waarin die zich voordoet, lijkt het mij legitiem het onderwijs te funderen in de functie van overdracht, in de brede zin des woords.

Het onderwijs is de institutie die bemiddelt in de overdracht van kennis, vaardigheden, levenservaring en competenties van de ene generatie op de andere. En dat permanent, en adaptief, en productief, zodat steeds optimaal wordt aangesloten bij de condities die de maatschappelijke context stelt. Enigszins pathetisch: de volgende generatie moet op de schouders van de vorige kunnen staan. De verschillen in het onderwijs zijn vergeleken met veertig jaar geleden wel groot, maar uitgerekend in deze kernfunctie is dat niet het geval. De praxis die voortvloeit uit deze kernfunctie (niet te verwarren met kerntaken) past zich aan de eisen van de tijd aan. Ik zou in dat verband willen spreken van *dynamische overdracht*.

Doelmatig onderwijs

Vanuit vorenstaande redenering is *doelmatig onderwijs de educatieve situatie waarin dynamische overdracht van kennis, vaardigheden, levenservaring en competenties optimaal wordt gerealiseerd, gegeven de condities en de beschikbare middelen*. Deze formulering van doelmatigheid veronderstelt een aantal uitgangspunten die aansluiten bij de door Webbink voorgestelde koers, maar die daar wel drie accenten in aanbrengen:

- duidelijkheid in cognitieve en sociale eindtermen – op dit punt kent het Nederlandse schoolsysteem een sterke traditie, die als uitgangspunt kan worden genomen voor het rekenschapmodel;
- nadrukkelijke sturing op organisatiestructuur van de school – optimale dynamische overdracht begint bij de leiding van de school: vanuit de organisatiestructuur naar de schoolcultuur naar het klassenklimaat (beleid rond zwakke scholen);
- versterking van de professionaliteit van leerkrachten door een grotere nadruk op de functie van dynamische overdracht door goed didactiekonderwijs, methodiekarticulatie en ‘leren van elkaar’.

Tot besluit

Het Nederlandse onderwijsstelsel heeft op het moment een slechte reputatie. Ik kan niet beoordelen of dit terecht is, ik wil daar ook geen oordeel over vellen. Maar ik veronderstel dat de zorg over de staat van het onderwijs direct samenhangt met de morele onzekerheid, zoals die keer op keer in bevolkingsonderzoek naar voren komt. Levert het onderwijs wel de kwaliteit die ‘onze kinderen’ nodig hebben om zich behoorlijk te

gedragen en zich ook nog staande te houden in de huidige samenleving? De ontplooiingsfilosofie van de afgelopen decennia heeft afgedaan, maar het is nog niet zo duidelijk wat daarvoor in de plaats is gekomen.

Er is een neiging om deze vraag zoveel mogelijk over te laten aan het veld, en daar zijn ook goede redenen voor. Toch kan over de kernfunctie van het onderwijs niet teveel worden gemarchandeerd. Op welke wijze vorm daaraan vorm wordt gegeven zal afhankelijk zijn van lokale omstandigheden, specifieke opvattingen en financiële middelen. Opbrengstgericht leren krijgt echter pas betekenis als we ons realiseren waarvoor we het doen. Over die vraag zal elke onderwijssituatie een permanent proces kennen van vragen stellen, idealen uiten en verantwoording afleggen. Dat geldt voor leerlingen, ouders, leerkrachten, directie, bestuur en samenleving. In die volgorde, en omgekeerd!

Literatuur

Kuhry, B. & De Kam, F. (2012). *Waar voor ons belastinggeld?* Den Haag: SCP.

3 Reactie Jaap Scheerens

Jaap Scheerens is hoogleraar aan de Universiteit van Twente

An essay on efficiency in education is badly needed in the Netherlands, where we have a lot of evaluations and assessments, but very few studies relating outcomes to (monetary) inputs. The conceptualization of efficiency as either less costly input with a stable level of output, or more/better output with fewer (or the same) input costs is quite helpful to structure the discussion. Within the education province the first interpretation (lower costs, similar output) is far less considered than the second, of improving outcomes, while the costs of bringing this about hardly ever enter the debate. The second interpretation (output efficiency) is quite prominent in recent educational policy with a strong focus of improving student outcomes in basic subjects. I will center my comments on three topics: ways to lower input costs that may not be detrimental to outcome levels, ways to improve outcomes that are relatively cost neutral and thirdly, some more specific comments on the way the author discusses accountability policies. Before that, I would like to raise two conceptual issues. Firstly, I wonder whether the concept of cost effectiveness, as, for example applied by Henry Levin, should not be used more explicitly in the paper. Cost effectiveness ratios, applied to education, compare inputs or treatments using gain in educational output as the denominator and the numerator is the cost associated with delivering the input or treatment in question. In this way alternative treatments (e.g. reducing class size compared to providing high quality text books) treatments can be compared on their effectiveness while taking costs into consideration. The second conceptual question I have is why “throughput” and malleable processes or “educational treatments” are not differentiated from inputs. Economic research concerning educational productivity has often concentrated on inputs like class size, teachers’ qualification, per student expenditure and teacher salaries, and not addressed process indicators like opportunity to learn, active teaching, and supportive organizational conditions at school and classroom level. Particularly when “output efficiency” is at stake it seems strange to give no place to “throughput” variables at school level that generally show higher effect sizes than the ones that are cited in Hanushek’s table in the paper.

Input efficiency: producing the same outputs with fewer inputs

I was surprised to find no proposals for increasing efficiency by spending less on certain issues. The reasoning put forward in the paper is that, given the high returns on human capital, cost reduction might work out as an example of being “penny wise pound foolish”. Closing off a potential area of innovative contributions from economists in education with a corny way of saying like this, is disappointing. In economist’s terms some high cost resource inputs in education, like class size have low elasticity, meaning that high increments on the input produce small or even negligible gains in output. Of course it would take some courage to draw attention to the likely absence of output loss when average class size would be increased by one or two pupils. If even economists are shying away from drawing attention to possible input inefficiencies, than whom else would be expected to do so? In the way the Dutch educational system is structured administration costs and managerial overhead deserve a second look, not to speak of the super inefficient way of bottom up school improvement, where millions are spent on re-inventing the wheel, nor of the billion EURO that is annually spent on special subsidies and an educational support structure that has never been evaluated for its effectiveness.

Output efficiency: producing more output with the same inputs

The author is quite pessimistic about the available knowledge on “what works” in education. His analysis appears to be based on a frame of reference that seems rather limited (two less successful intervention studies, mentioned on page 10). Quite in line with the blind spot concerning malleable process factors in education at school level, as noted in the above, the author ignores studies in educational effectiveness and the knowledge base that has been established in this field. Authors like Walberg and Hattie, have produced reviews, based on hundreds of meta-analyses (Walberg, 1985, Hattie, 2009). Increasingly these meta-analyses are not just based

on observational studies, but draw more and more also on randomized experiments, of the kind that the author is (rightly) proposing as a constructive way ahead. Robust factors from educational effectiveness research have been used for evidence based programs, in the United States and elsewhere, (known under the heading of Comprehensive School Reform programs), and are showing modest but educationally relevant results (cf Borman et al. 2003). Educational innovation policy in the Netherlands would gain in efficiency if such type of programs were used more often.

The author is almost as prudent and cautious as it comes to his assessment of what we know to improve educational performance, as he showed himself with respect to seeing opportunity for efficiency gain on the basis of input reduction. His caution on “what works” research is warranted to a degree, as indeed, most results and meta-analyses still depend on observational studies, and results of meta-analyses show highly diverging mean effect sizes (Scheerens, 2012). Yet, levers for performance improvement exist, not just as system level measures, such as school choice and accountability policies, but also at school and classroom level. Different capacities of schools and teachers to provide opportunity to learn, to make good use of performance feedback and provide “focused” teaching, based on enacted “pedagogical content knowledge” of teachers are some examples of promising strategies.

Smart accountability policies and a learning strategy for educational policy

Smart accountability policies, according to Dinand Webbink, should prevent negative side effects, such as strategic behavior by schools, by means of control from the Inspectorate, in preventing schools to display this behavior. It is questionable whether more rigorous external control would solve the problem. With regards to the Netherlands, external school accountability exists, but is less “high stakes” than in the many examples that we get from the USA. In a recent presentation Helen Ladd spoke favorable about the mixed arrangements for school accountability in the Netherlands, including examination and test results, but connected to school inspection, in comparison to high stakes testing in the No Child Left Behind program (Ladd, 2012). More generally, combination of external and “internal” school accountability, (cf Carnoy et al., 2003), including formative assessment, might be more effective than just more external control.

Step wise policy formation, accompanied by rigorous research, is a very good recommendation. In order to learn most from such a strategy, it would be important that even the initial trials could benefit from already available evidence, in other words, that explicit programs would be tried out as treatments. Otherwise, true experiments evaluating ad hoc and “black box” programs would not produce much learning for educational policy. Unfortunately this strategy does not combine well with the predominant bottom up school improvement philosophy in the Netherlands: The evidence based way of policy formation, the approach of “reforms as experiments” is hard to reconcile with a development strategy in which there is no place for externally shaped “treatments”.

References

- Borman, G.D., Hewes, G., Overman, L.T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125-230.
- Carnoy, M., Elmore, R. and Siskin, L., (2003) *The new Accountability. High Schools and High-Stakes Testing*. New York and London: Routledge Falmer.
- Hattie, J., (2009) *Visible Learning*. New York: Routledge.
- Ladd, H. (2012). *School accountability and monitoring systems: Insights from the U.S. and other countries*. Key-note address at the Improving Education through Accountability and Evaluation – conference. Rome, 3-5 October.
- Scheerens, J. *Theories on educational effectiveness and ineffectiveness*. Key-note address at the EARLI- special Interest Group on Educational Effectiveness conference. Zürich, 29-31 August.
- Walberg, H.J. (1984). Improving the productivity of American schools. *Educational Leadership*, 41, 19-27.
- Webbink, D., (2012). Efficiency in Education. Essay prepared for the Education Council. Rotterdam: Erasmus School of Economics.

4 Reactie Gert Biesta: Educational Efficiency, Educational Aims, and the Complexity of Educational Systems

Gert Biesta is hoogleraar Educational Theory and Practice aan de Universiteit van Luxemburg⁶

The question of the efficiency of education is an important one, not in the least because modern societies tend to invest a significant amount of public funding in their education systems. In a time of austerity it becomes even more important to ask whether such funds are deployed in the most efficient way possible. It is for this reason that a discussion on the topic is timely, and the essay by Webbink provides a useful and succinct starting point for this. What is particularly helpful is that it consistently works through insights from the field of educational economics. In what follows I will look at the question of educational efficiency from the perspective of educational theory. Doing so will allow me to indicate some of the limitations of the economic perspective – limitations that, as I will suggest, particularly surface where it concerns the implications for education policy.

Distinguishing Educational Efficiency from Educational Effectiveness

The first point to make is that the question of educational efficiency should be distinguished explicitly from the question of educational effectiveness. Educational effectiveness concerns *whether* a particular educational approach or strategy will bring about the desired results or outcomes and also whether one particular approach or strategy will do this more securely than another. The question whether the study of literature is a better way to promote morality than, say the modification of behaviour through rewards and punishment is, therefore, a question of educational effectiveness. The question of educational efficiency, on the other hand, does not focus on the 'whether' but on the 'how' and asks about the resources that need to be deployed in order to bring about a desired result or outcome.⁷ More specifically it is interested in identifying the minimal amount of resources needed to bring about a particular result or outcome. Whether an incandescent light bulb provides better street lighting than a LED light bulb is, therefore, a question of *effectiveness*, whereas the fact that the LED light bulb uses far less energy than the incandescent light bulb concerns the question of *efficiency* – which also shows that the most efficient way is not necessarily the most effective way, as we may come to the conclusion that the incandescent light bulb does a better job at lighting the street.

This means that I agree with Webbink's definition of efficiency as being about "the optimal use of resources" and that, in this regard, his distinction between "output efficiency" and "input efficiency" is a useful one. I also agree with his observation that education is normally concerned with multiple outputs (see below), so that a strive for efficiency in education always needs to take into consideration how an increase in certain 'outputs' impacts on the achievement of other 'outputs' that are considered to be valuable or desirable. In this sense we might say that the question of educational efficiency is an entirely *technical* question as it simply is interested in the optimal deployment and utilisation of resources. This means that, strictly speaking, it is also a relatively *limited* question, because it is a question about the *formal* aspects of system performance. Education practice and policy are, however, never just about formal questions but always have to engage with substantive questions as well. When we look at the question of educational efficiency from that angle, things become a bit more complicated. In the next sections I would like to make three brief observations, one concerning the 'outcomes' of education, one about the process of education, and one about the means of education. In each case I will expand Webbink's economic perspective by raising what I see as key educational questions. In a final step I will show how this leads to a rather different approach to the question of implications for policy.

⁶ From 1 January onwards: Faculty of Language and Literature, Humanities, Arts and Education, University of Luxembourg, Route de Diekirch, BP2, L-7220 Walferdange, Luxembourg. e-mail: gert.biesta@uni.lu

⁷ While Webbink generally sticks to the question of efficiency, in a few places he seems to conflate it with the question of educational effectiveness, for example in discussing experimental and quasi-experimental designs in efficiency research which, in my view, are actually designs for effectiveness research.

The Question of Educational Aims and Purposes

In one sense we could say that Webbink sticks to his trade by just asking formal questions about the efficiency of the educational system and by highlighting that the main problems we encounter when trying to answer such questions are technical problems. This is why he states a number of times that the "key problem" in the discussion on educational efficiency is the problem of *measurement*, that is, of how to measure "all the relevant input factors and outcomes" (p.7). Notwithstanding this, he does articulate a very particular (and in my view narrow and problematic; see below) view about what education is supposed to be for, in that he repeatedly refers to 'student performance' – often to be operationalised through test scores – as one of the main relevant 'outcomes' of education, and also to the production of human capital as a main aim of education. In my view, however, one of the key challenges in the discussion about educational efficiency is not the technical question of measurement but first of all the substantive question about which 'outcomes' are considered to be desirable. (I will explain the reasons for putting the word 'outcomes' in quotation marks below.) While economists might see the main (or even only) function of the education system as that of the production of human capital, from an educational perspective this is only one function to which the educational system might be orientated.

Elsewhere (see Biesta 2010a; see also Biesta 2007a) I have suggested that education always functions and always ought to function in relation to three different domains, the domain of *qualification* (the transmission and acquisition of knowledge and skills); the domain of *socialisation* (the encounter and engagement with existing cultures, traditions and practices); and the domain of *subjectification* (the formation of the person, for example orientated towards critical autonomy, rationality and morality). The problem is that many measures of student performance that are being used to assess both the effectiveness and efficiency of educational systems take an extremely selective and hence extremely narrow view of what it is that education ought to achieve – often only focusing on test scores in a small number of subjects and thus only focusing on the qualification function of education. (This is, for example, the problematic way in which the 'quality' of education is operationalised in PISA.) While in recent years there has been a growing interest in the socialisation dimension of education – for example through such 'educations' as environmental education, personal, moral, social and health education, or citizenship education – the subjectification dimension is still relatively absent in discussions about what good education is and what education should aim to achieve. Yet the latter dimension is absolutely crucial if we see education not just in economic or economic terms, but also as an investment in the quality of social and democratic life more generally. Such 'outcomes' are, of course, far more difficult to measure and assess, but they belong to the very heart of the justification for a *public* education system.⁸

Causality or Complexity: On the Nature of Educational Systems

That education always also needs to be orientated towards subjectification – that is, towards the becoming-subject of children and young people – also implies that we should never just think of the student as an object of educational interventions, but always also as a subject of action and responsibilities in his or her own right. This means that education always needs to engage with the double task of engaging children and young people with the world and emancipating them from that world at the very same time (see Meirieu 2008, p.91). After Auschwitz – to use Theodor Adorno's phrase (Adorno, 1971) – education can never just be about the adjustment of children and young people to the existing socio-political order but always needs to work towards them being able to make up their own minds and think and speak for themselves. In my view this reveals another shortcoming in the way in which Webbink discusses the question of educational efficiency, because he depicts education entirely in terms of a (quasi-)causal production process in which the student is only the object of educational 'inputs' and the producer of educational 'outputs' but never a thinking and acting subject in his or her own right. Such a (quasi-)causal conception of the educational process – which also underlies much

⁸ This does, of course, raise the question of the increasing privatisation of education – something that is more prominent in some countries than in others – where education is increasingly being repositioned as a private good and hence an investment in personal advantage, rather than a public good that is always also orientated to wider social and political values of freedom, equality, and solidarity. From this perspective Webbink's reference to the way in which schooling results in an increase in 'earning power' reveals either an orientation towards education as a private good, or a lack of awareness of the importance of the difference between education as a private good and education as a public-democratic good.

of the work on educational effectiveness (see Biesta 2007b; 2010b) – is based on a fundamental misunderstanding of the dynamics of educational processes and practices. One way to make clear why this is so, is through the language of systems theory (see also Osberg & Biesta 2010).

Referring to education in terms of 'production' – and more specifically to the efficient production of certain 'outputs' as a result of certain 'inputs' – suggests a view of education as (ideally) a deterministic system where, once we have sufficient knowledge about 'what works' and have sufficient control over all the relevant variables, we can establish secure connections between inputs and outputs. While such expectations may be realistic in the case of closed deterministic systems – that is, systems where every input produces a predictable outcome and where there is no possibility for interference from the outside – education systems are systems that at most work probabilistically but never deterministically. The reason for this lies in the fact that education systems are open, semiotic and recursive systems. They are *open* in that the boundaries with the environment are never completely closed – children, for example, go home after school. They are *semiotic* in that the interactions between teachers and students are not based on physical push and pull but on meaning and interpretation – students trying to make sense of what teachers are saying and doing and, through this, learn from their teachers. And they are *recursive* in that the actions of the 'elements' in the system (i.e., teachers and student) feed back into the system and alter the direction in which the system will develop, and the main reason for this lies in the fact that the 'elements' in the system are not stimulus-response machines but thinking and feeling beings who, based on their perceptions and interpretations, can choose to act in a range of different ways.

To think of education just in terms of production – and the production metaphor figures prominently in Webbink's account – overlooks the specific nature of educational processes and practices and instead works with an ideal image that bears no resemblance to how education actually 'works.' By putting the complexities of educational processes and practices in a 'black box' in between inputs and outcomes, Webbink thus misses crucial opportunities for analysing and understanding how connections between inputs and outcomes are actually achieved and thus for understanding what it might mean to increase the efficiency of what is going on. Before I indicate what this implies for how we think about the policy implications, there is one more comment I wish to make about efficiency, which has to do with the *non-neutrality* of educational means.

The Non-Neutrality of Educational Means

The argument for educational efficiency as presented by Webbink centres on the idea of identifying those actions and interventions that result in the most efficient production of certain educational outcomes. While at a general level and in formal terms this is indeed a correct way to think about efficiency, this way of thinking encounters a problem when applied to the field of education. The particular problem here has to do with the fact that the means that are used in education – that is, the teaching strategies and approaches, but also many other input variables that can be controlled by teachers or the education system more generally – are never neutral with regard to the ends they aim to achieve. This is a well-known problem in the discussion about educational effectiveness, because whereas it might be possible to identify effective ways to bring about desired 'outcomes,' the most effective way is never automatically also the most *educational* way. The standard example here is that of punishment (see Carr 1992).

While punishment may be an effective way to promote certain behaviours and discourage others, and while research might even show us that it is the most effective way, we may still decide not to use punishment in education. The reason for this is an educational one and has to do with the fact that the use of punishment in a classroom is not a neutral means to bring about a particular result, but at the very same time teaches students something (for example, that in some cases, force can be justifiable). The issue here, in more general terms, is that students never just learn from what teachers say and do but also from how they say it and how they do it – and most students are actually extremely good in spotting the contradictions between the two. This therefore puts an important educational 'boundary' on any discussion about efficiency, because even if we were able to identify the *most efficient* way to bring about a particular educational 'outcome' there is still the question whether this is also the *most educational* way of proceeding. While it may be more efficient, for example, to give

each student their own laptop and let them work individually on tasks, we may have good educational reasons for letting them share laptops. While this may slow down the process, it provides crucial opportunities for students to learn what it means to collaborate and this might be a good educational reason for using a less efficient (but educationally more worthwhile) approach.

Implications for Policy: Pressure or Professional Responsibility?

This brings me to my final point, which has to do with the implications for policy. Webbink formulates his suggestions for policy on the basis of his 'black box' conception of education as a process where certain inputs through certain (quasi-)causal processes are supposed to result in certain outputs. I have suggested that such a conception may make sense with regard to deterministic physical systems, but that it does not make sense in the case of social systems such as education, since education is an open, semiotic, recursive system. Looking at education in this way one could of course ask how it is possible in education to have any connections between 'inputs' and 'outcomes' at all. The answer to this question, when looked at in terms of systems theory, lies in a principle that I have referred elsewhere to as 'complexity reduction' (see Biesta 2010c).

The idea here is that in order to move open, semiotic recursive systems into the direction of closed deterministic systems we need to reduce the number of available 'options' for the 'elements' in the system. (Fast food restaurants are a 'good' example of this.) Concretely this means that we can make things 'work' in education when we reduce [1] the openness of the system, that is, its interaction with the outside world (and we do this, for example, by placing children in school buildings, and by giving them a curriculum); when we reduces [2] the opportunities for interpretation (which we do, for example, through assessment which is a process in which we select which of the interpretations our students generate are 'valid,' 'right,' 'correct,' etcetera); and when we reduce [3] the recursivity of the system (which ultimately can only be done by stopping the 'elements' from thinking, judging and acting). In this way the unpredictability of the system can radically be reduced and education can begin to operate as a much more deterministic system.

While this is partly where Webbink seems to want to 'push' educational systems – and his argument here seems to be both one of effectiveness and of efficiency – looking at it through the lens of systems theory makes it possible to see at *what price* we might improve the effectiveness and efficiency of educational systems. While *to a certain extent* this is a price we are willing to pay in order to make education work, we can see that to 'push' education in the direction of a deterministic mode of operation ultimately requires that we take out interpretation, thinking, and reflection. It thus requires that we treat students and teachers as objects – as robots, we might say – rather than as subjects. This, as I have suggested, is a price we should *not* be willing to pay. It is here that I am concerned about several of Webbink's suggestions for policy, as they tend to be formulated in a language of force – he refers to 'putting pressure' on schools and teachers, for example – that seem to want to take out interpretation, reflection and, most importantly, responsibility.⁹

Such a 'top-down' approach to educational policy, even if the ambition is to increase the efficiency of what admittedly is one of the most costly endeavours modern societies are involved in, thinks of those who make the educational system 'work' – teachers and also students – as just objects of intervention and control, rather than committed professionals who work in complex settings towards a broad range of educational aims and ends. It is therefore also in terms of policy implications that the production conception of education is misleading and misguided and that an understanding of education as a complex, open, semiotic and recursive system opens up a more realistic and more promising window on how policy might engage with question of efficiency in a way that acknowledges the specific nature of educational processes and practices.

⁹ I also find the suggestions about competition problematic. First of all, after the recent banking crisis I find it difficult to believe that economists would still maintain "that market competition leads to efficient outcomes" (p.11). But in many if not all cases it is also an illusion to think that competition through school choice can "stimulate schools to improve the quality of their education" (p.11). The main reason for this is that in many contexts it is simply too expensive to provide any meaningful choice for parents or students, as many (smaller) communities can only support a limited number of schools.

Conclusion

In my reflections I have suggested that we should not understand education as a (quasi-)causal input-output system of production but rather as an open and complex social system, as system that functions because of human communication and interpretation. Looking at education in this way highlights different avenues for increasing the efficiency of the system. Rather than putting pressure on alleged input-output chains, my approach suggests that efficiency might be increased through a reduction of the complexity of the system, that is, through reducing openness, semiosis (interpretation) and recursivity (reflection, judgement and action). Looking at the question of educational efficiency in this way then makes it possible to see that a reduction of the complexity of the system's functioning always comes at a price and that there may well be a critical tipping point where the education system ceases to be educational, and becomes a system of control rather than emancipation. That is why the question of educational efficiency is ultimately an educational question rather than one that can entirely be captured by economic frameworks.

References

- Adorno, Th.W. (1971). *Erziehung nach Auschwitz*. In Th.W. Adorno, *Erziehung zur Mündigkeit. Vorträge und Gespräche mit Hellmut Becker 1959-1969* (88-104). Frankfurt a.M.: Suhrkamp.
- Biesta, G.J.J. (2007a). De burgerschool als leerplaats voor democratie. In Onderwijsraad, *Alternatieven voor de school* (13-30). Den Haag: Onderwijsraad.
- Biesta, G.J.J. (2007b). Why 'what works' won't work. Evidence-based practice and the democratic deficit of educational research. *Educational Theory* 57(1), 1-22.
- Biesta, G.J.J. (2010a). *Good education in an age of measurement: Ethics, politics, democracy*. Boulder, Co: Paradigm Publishers.
- Biesta, G.J.J. (2010b). Why 'what works' still won't work. From evidence-based education to value-based education. *Studies in Philosophy and Education* 29(5), 491-503.
- Biesta, G.J.J. (2010c). Five theses on complexity reduction and its politics. In D.C. Osberg & G.J.J. Biesta (Eds). *Complexity theory and the politics of education* (5-13). Rotterdam: Sense Publishers.
- Carr, D. (1992). Practical enquiry, values and the problem of educational theory. *Oxford Review of Education* 18(3), 241-251.
- Meirieu, P. (2008). *Pédagogie: Le devoir de résister. 2^e édition*. Issy-les-Moulineaux: ESF.
- Osberg, D.C. & Biesta, G.J.J. (eds) (2010). *Complexity theory and the politics of education*. Rotterdam: Sense Publishers.

5 Reactie John Hattie

John Hattie is hoogleraar onderwijs aan de Universiteit van Melbourne

Efficiency in Education

Let me start backwards. There are two Policy implications and challenges.

1) Smart accountability policies

This is a well travelled route and while there is evidence that having smart (clear and transparent) targets and accountability there is a tipping point – most Western countries have such systems but there is evidence of short one-off gains (as teachers learn to teach to the test, narrow the curricula, etc.) there is no evidence of sustained improvement from these systems. Yes, stopping exclusions from the test is another one-off loss (scores go down) but as many countries have shown this can hover around 3-6% even if only due to sickness – nothing much then changes.

My argument is that most accountability solutions are based on the wrong premise for a theory of change. Most often they are based on “shoving thermometers” into students and then

- Making claims about relative achievement. In most systems these relativities are closely tied to real estate prices around a school; they are based on narrow assessment more akin to IQ measures (most IQ measures include literacy (esp. vocabulary) and numeracy), and there is evidence therefore of “flat lining” – as it is very difficult to change the average IQ of a nation.
- Ignore what schools are charged to do – add value. My claim is that accountability should be premised on the notion that “every student deserves at least a year’s growth for a year’s input”. I know from the Visible Learning synthesis that this is $d > .40$ per child (as a starting ambit claim), and while some students need much more than 1 year’s gain per year, every student should at least be shown to markedly make changes over a year. Most accountability systems ignore this, or make gross estimates (AYP in US), such that teachers cannot see how and where they have made these gains.
- Ignore a fundamental problem – teacher quality varies WITHIN schools and this variance is much larger than between schools. The greatest variability comes from teachers not sharing a common conception of progress.
- If we demand teachers and schools enhance students we should support them with an assessment system that helps them know their impact – on a regular basis, across all students. The Netherlands have been leading the world in many of these systems. The system I created for NZ is based on these above notions – teachers and schools can regularly access the national assessment tool (e-asTTle), and get immediate reports on how they have impacted, the magnitude, who, and about what. It also helps dramatically in ensuring teachers across the nation have common conceptions of progress, and then can be used to individually create targets. Most important, it leads to two major implication: (a) teachers can discuss (safely) their impact on students with other teachers and school leaders, and (b) it leads to implementing a model of “student assessment capabilities” (Absolum et al., 2008) – and thus both students and teachers are working, understanding, and interpreting the assessment information to make progress.

Thus, be careful about global statements re accountability. Measurement experts, parents see this as “more tests” whereas my recommendation is to see this as “more systematic interpretations to assist in progress” and knowing one’s impact.

2) A learning strategy for educational policy

It is clear to me that systems that adopt an “improvement” mind set. This is contrary to the accountability, the choice, the high achievement mind set. Given that each year a new cohort of students start each year – again

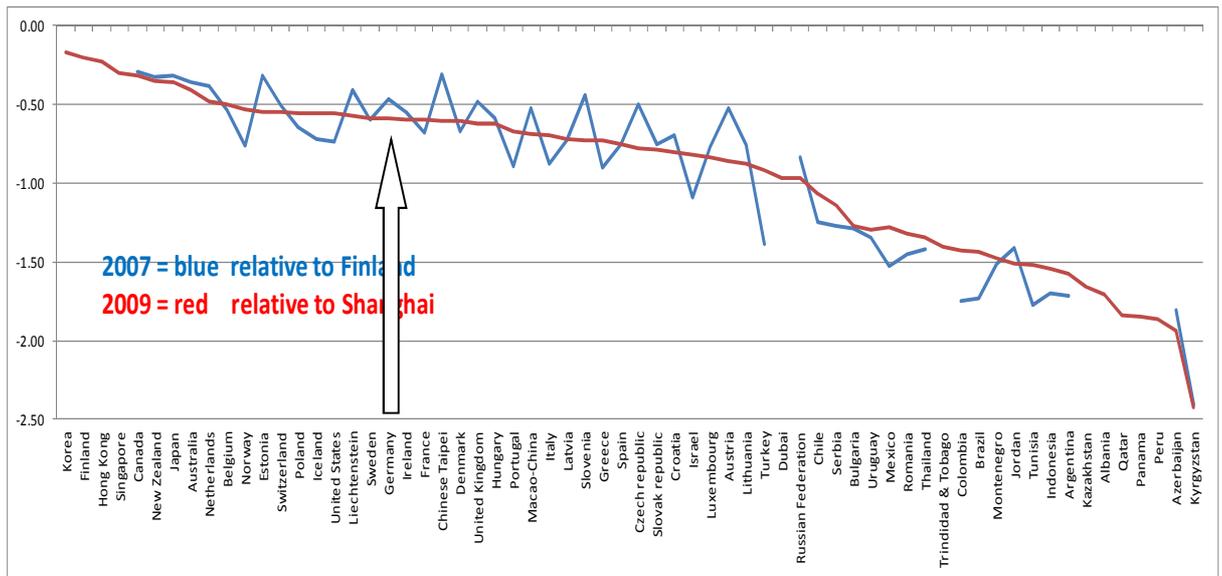
for teachers – means they live in constant improvement scenarios. So how can a system become an “improvement” system – finding ways to constantly support school leaders and teachers. As critical, how does a system become an improvement system and learn from success and failure – most systems work on an “election cycle” mentality, reinvent old wine in new bottles, worry about the structure and peripherals, and fail in the evaluation of their own programs. I understand why this is so hard in democracies and transparency – how can a system admit its funding choices have not worked. I have tried to find systems that allow for independent evaluation without biting the hand that feeds it – there are few. NSW is trying an independent evaluation system chaired by a nationally (non-NSW) esteemed academic and this offers a way forward.

There IS a practice of teaching. My Visible Learning synthesis shows that this is so – we know probabilities of success of many interventions. We know what has low probability, and what has high probability. The issue is why do not systems and teachers believe this, what do they continue to allow schools to trail regularly failed programs (in the name of autonomy) but most important why do we not demand evidence from schools about the impact of their programs. Why do we not demand a flavour of collective impact in a school?

The concept in this paper is “efficiency” but it is phrased in terms of production functions. Another (too often absent) notion is efficacy in time – Imagine tow similar groups of students – one teacher moves these students to a target in much less time than another. This is efficiency. Take as another example, acceleration. Among the most top in the VL rankings – but when implemented by having students skip a year, principals lose a year’s funding. All the contingencies in our schooling system are weighted against efficiency. Where is the debate about efficiency in learning, in schools, where are the rewards for efficiency.

It thus makes sense, given the lack of efficiency in our system that:

- 99% of teachers receive satisfactory ratings each year - it is clear we use the “can you improve achievement” benchmark, not the “can you markedly improve achievement” benchmark?
- When we rate all teachers as good or great, those who are truly exceptional cannot be readily identified.
- When beginning teachers believe they also deserve the highest performance ratings, when they (within 6 months) prefer they own class and work alone, then it is unlikely we will identify excellent in teaching.
- When about half of what teachers learn about teaching is learnt in the first year, and half as much again in the second year, the trajectory of teaching expertise beyond 3-5 years is flat – no wonder it is hard to identify expertise, no wonder most performance pay systems change nothing much.
- In the last 20 years in the US the student population has grown by 17%, the number of teachers by 32%, and the non-teaching staff by 46% -with no evidence of a change in the impact on students. We do not value expertise or efficiency, we value bulk! It the Netherlands is to rise to the top of the world it needs an average improvement of .50 sd. This is a huge effect size change; about an increase of over one year for each student from where they are now. This is based on the average effect-size from Finland or Shanghai in Reading, Math and Science.



Reaction to the review

Production functions are powerful in education, but more recently value-added models have made a great contribution to the debates about efficiency and effectiveness. These are barely mentioned and this makes the paper dated.

The sections on money ignore the move about 10-20 years ago from “Does money matter” to “How money matters” (e.g., See Norton Grubb’s synthesis in his recent book). In VL, I show that across four meta-analyses the average effect-size from the influences of differential finances is .23 (based on over 3m students), and ranked 99th out of 130 influences. The comments were:

Although the meta-analyses seem to indicate that money does not seem to matter, this would be a misleading conclusion. Childs and Shakeshaft (1986) undertook a meta-analysis of studies on the relationship between educational expenditure and student achievement and showed that there was a minimal relationship between the two, and the most positive relationship related directly to the costs of instruction; for example for teacher salaries and instructional supplies. Teacher salaries, in turn, were more related to years of teaching experience and not teacher quality. Rolle (2004) also argued that more money was not necessarily needed but that there should be more productive use of existing resources. This is consistent with the claims often made by Hanushek (1989) that there is no consistent statistical relation between educational expenditure and measures of student performance. For example, Hanushek (2003) correlated secondary school spending per pupil and mathematics scores across 23 countries (from TIMSS, 1998) and found a correlation of $r = 0.06$. So often money is added into the education system with little attention to the efficiency or effectiveness of education outcomes. It is not the amount of money spent that is important, but how it is spent.

Murdock (1987) reported financial aid on college students had a small but positive effect on student persistence, and enabled low-income students to persist at a similar rate to that of students from other socioeconomic groups. While student financial aid is an important tool in helping students to stay in college, other factors such as the type of institution, the length of course, and year in which aid is provided (effects are greater with more senior students than with first-year students) all mediate the effects of financial aid.

In a rebuttal to the claims about the limited effect of increased finances, Hedges, Laine and Greenwald (Greenwald, Hedges, & Laine, 1996; 1994) analyzed the effects of differential school inputs on student outcomes. Their analysis showed systematic, positive patterns in the relations between educational resource inputs and student outcomes. An increase in per pupil expenditure of \$500 increased the effect on achievement by $d = 0.15$ for per-pupil expenditure, $d = 0.22$ for teacher education, $d = 0.18$ for teacher

experience, $d = 0.16$ for teacher salary, and $d = 0.04$ for teacher/student ratio. Thus we can expect “comparable and substantial increases in achievement if resources were targeted to selecting (or retaining) more educated or more experienced teachers” (Greenwald et al., 1996, p. 380). There is little evidence, however, to justify the notion of “substantial” but there is much consistency with other meta-analyses about the importance of the teacher (and costs associated with enhancing teaching).

The seemingly limited effect on finances can be related to that fact that (a) most studies have been conducted in well resourced countries (e.g., the United States, United Kingdom) where the variance in resources to schools is not so substantial; (b) most finances in schools are tied up not in discretionary but in fixed costs (such as teacher’s salaries, busing, and buildings) and these do not vary in proportion of costs across schools within any country; and (c) if the school composition effects are much greater at the within- than between-school level, then costs could make a difference within schools and less of a difference between schools (and most are currently focused on between and not within schools). As Hanushek (2002; 2003) has long argued, there are few incentives for a teacher to maximize achievement as most of the financial incentives are related more directly to school rather than teacher differences.

The emphasis may need to be *not* on the notion of “Does money make a difference?” but on “How does money make a difference, particularly above and beyond the fixed costs of running a school (capital, lighting, salaries)?” It is difficult to imagine money does not make a difference at these critical margins. Jonathon Kozol (2005), in his scathing analysis of the restoration of apartheid schooling in America, *The Shame of the Nation*, makes the claim that he will believe money is not important when wealthy parents stop spending so much money on their children’s educations. He cites Deborah Meier comment that “I’ll believe money doesn’t count the day the rich stop spending so much money on their own children” (p. 59).

One of my Phd students has recently submitted her thesis on HOW money influences and there is some (not strong) evidence that where more is spend on instructional compared to structural resources in a school there is evidence of achievement improvement. Yes, as the paper comments – we seem not able to convert extra money into better results – as noted in the Widget effect we spend it on more adults, not necessarily more expertise in learning. (Take for example teaching assistants, Blatchford et al., have clearly shown that they have a zero to negative effect especially on struggling students and the investment should be spent on up skilling these assistants, if at all.)

Choice and Autonomy (I have taken some of the below from my submission I am preparing for a local enquiry).

The promise of school autonomy or decentralisation of decision making powers is indeed a current debate – ironically in the same times as state and federal governments assume even more control over the schooling debates. Autonomy has attained the status of a desirable and must-have, is promised by governments to improve the quality of schooling, and the claim is that decision making is best made locally. In my context, “A key direction of government policy in Victorian Government schools over the past two decades has been increased devolution or decision making to the school level”. School autonomy can be understood as the delegation of certain tasks from one level of governance to another (in this case centrally to schools).

What are the various meanings of autonomy in the governance of schooling

Autonomy is a polymorphous concept: it has many and varied meanings. Further, it is often used as a “desirable attribute” but its effects depend on *which* aspect or variant of autonomy is enacted and what notion of ‘desirable’ is used. The move to autonomy is also mostly coupled with many others current “desirable attributes” claims, such as choice, competition, and accountability. There is a belief that by creating choice and competition and then locating responsibility at the local level there will be incentives to improve the quality and accountability of each school. Often autonomy is contrasted with central control (as if decision making is one or the other), and this central control is The alternative is seen as not sustainable: that is, if all schools are treated similarly by government then bad schools can continue to exist and parents do not then have information and choices to move to better schools.

The argument is that schools, like businesses, should thrive or not based on their proficiency to meet consumer demands or to create consumer demands. This leads to providing autonomy to local schools and thus enhancing choice on the demand side, provide suppliers and consumers with more information (e.g., My School websites), and the intention is that consumers will drive out the bad and move to the good schools. Schools will thus quickly learn to change their behaviours and practices to move towards being “a good school” and thus staying open and successful in the eyes of parents. To make these changes, it is argued, schools must have autonomy.

The presumption under these ‘autonomy’ arguments is that there is much information available to inform consumers and there is a need for accountability. Too often, the notion of accountability is thus introduced for the purpose of providing standardised comparable outcomes (student test scores). There is already much other information provided by schools - as school marketing research demonstrates schools tend to provide information about the peripherals of schooling – the buildings, the students, the library, the sports facilities, the ICT available, the size of classes, medals, and exam results. It is rare to see marketing based on schools having excellent teachers – thus the greatest source of variance in quality is missing. The more recent moves is to demand that many correlates of good teaching are offered - especially standardised tests - but much of this test scores information is limited by a) being constrained to the easiest to measure skills, b) often these test scores are no more than generalised ability measures, c) they cover a very limited part of the curriculum that parents may value, and d) it is difficult to ascribe causation to these test schools to teachers (e.g., in Australia NAPLAN average school means are more measures of house prices – especially as they do not measure growth).

Autonomy is given and taken away.

Before venturing to the details of autonomy it is important to talk about control. Because autonomy is given and taken away as a function of who controls what. There are many types of control in school systems (Hattie, 1992; 2012; Hannaway, 1990) and these have been subject to multiple pendulum swings over the past decades.

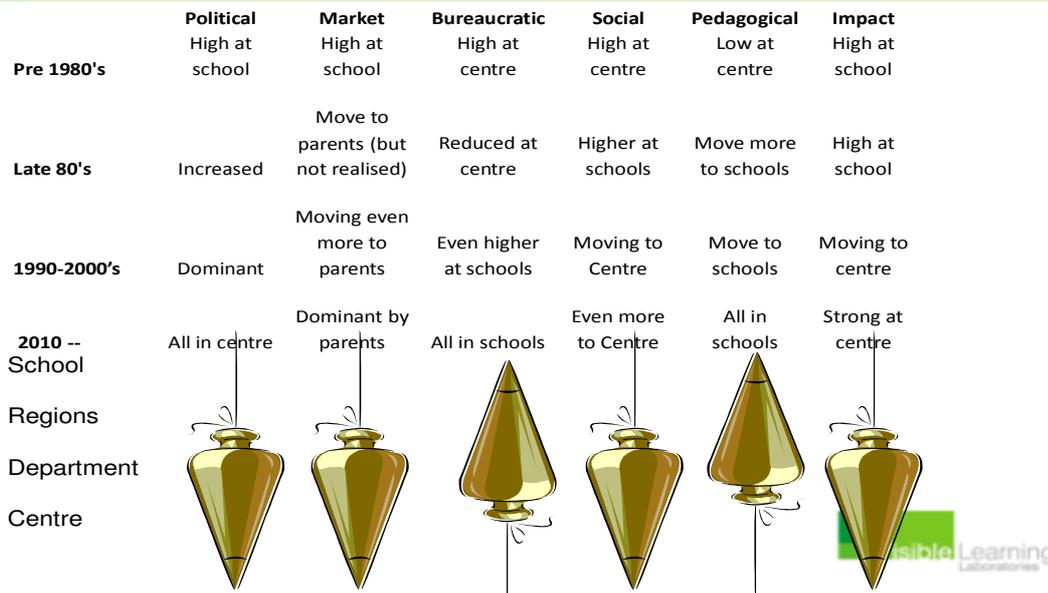
1. Political control, or the action by elected officials. The elected officials typically decide "who benefits" and "who pays". This form of control has increased substantially over the past five years at the Federal level. We moved in the 80's from the 'old' Department of Education which used to be independent and advise the Minister, to the model of a 'new' Ministry of Education whose personnel are to carry out the executive orders of the Minister. This has led to the current practice whereby Ministry officials enact the Ministers' policy, a lack of critical comment and debate by state departments, preaching of the Ministerial policies, and a protection of the centre from the masses (the masses in the schools and in the community). In the past decade, this has become even stronger as Federal initiatives (and funding) have begun to rule the debates – and we have the era of National curricula, National standards, and National testing that are dominating the debates. The irony is that the independence and autonomy has been taken from schools through to Ministries and now Federal agencies – and the new promise is that if you behave we might give you some of this autonomy back.
2. Market control or control through consumer or client choice, is demonstrated by the increased involvement of parents in schools. This attention to parents is becoming a major source of debate and control, and the ubiquitous notions of “choice” are peeling in the school debates (Hattie, 2012). The major fault of market control is that uninformed consumers are poor regulators. In the education enterprise it is a daunting task to become an informed consumer: the results of education take years to realise, the outputs are difficult to define and measure, the costs of switching schools are usually only available to the more well off (this is why vouchers and ‘choice’ will primarily swing more parents to send their children to non-Government schools), and uncertainty leads to over emphasizing incorrect (and usually easier) indicators. Market control presumes that children are equally advantaged because their parents are equally informed: the converse is ethically indefensible: why should children suffer because their parents are uninformed.
3. It is unfair to allow too much parent/local control over schools. This leads to greater inequalities as schools in well resourced areas have greater parental capital to draw upon and those in low resourced areas struggle to obtain such capital. This inevitably leads to greater dispersion, and it must be asked why students should be punished because of their parents reduced know-how to contribute to the governance

of schools. This is not dissimilar from Hanushek, Link & Woessman (2011) finding that “autonomy reforms improve student achievement in developed countries, but undermine it in developing countries... (where) it hurts student outcomes” (p. 25). The same findings apply within a country as well. In developed countries the most benefits came from autonomy over academic content to teach within a curriculum, some effects or personnel autonomy, and less for budgetary autonomy.

4. Bureaucratic control considers that there are tasks to be done in order for an organisation to perform productively, and this form of control has been most decentralised since the 1980's. Central agencies used to have forms, formula, and standard answers to distribute to all schools and this has well gone. This increased decentralisation of bureaucratic control to schools appears to be welcomed by school leaders (although they complain loudly) as it has helped redefine their jobs as project managers. When these now local decisions are taken away from schools the complaints are vociferous.
5. Social control depends on agreed goals. The moves by governments to insist a vision statements, purpose and school development plans are an attempt to exert social control. This form of control assumes that all are committed to the goals, and we will behave as best as we can to promote these goals. The problem is that developing a productive culture among teachers is much more difficult and much trickier than promulgating directives. Most schools end up with school plans that differ at the margins from other schools as the business of education is not that different across contexts – that is, Australia has very low between school variance (and much higher within school variance). The differences are more within than between schools.
6. Curricula and Program control. The business about what programs and curricula a school can offer has mixed messages. Curricula are becoming more national, but the form of particular programs in which this curriculum is wrapped is more local. How the curricula are presented, the timetabling, the allocation of teachers to programs is very much locally controlled. The major deficiency is the low levels of learning across schools about the optimal ways of making these decisions.
7. Pedagogical control. We have moved from very similar looking classrooms, straight lines of desks, chalk and talk, through to today where schools adopted “fads” such as open classrooms, discovery learning, constructivism, learning styles and so on (see Hattie, 2009b for evidence of their low effects on achievement). Cuban (1993) has demonstrated that teaching has barely changed over the past 150 years, and the ‘grammar of schooling’ as he terms it relentlessly gets imprinted on each new innovation. While there has been many evidence based methods of teaching, this does not mean that they have been implemented as there is resistance both to changing current methods and from the fundamental belief in the “autonomy” of teachers to choose their methods. It is unlikely that school systems will intervene to prescribe optimal teaching methods – although the most dramatic improvements in the US system have occurred when this has happened (e.g., Chicago, Guilford County, SEMOE and NCEA in NZ). It is the case, however, that often these pedagogical imperatives are mandated in systems with very low performance – they create a common debate, a collaborative implementation, and an attention to fidelity of implementation (but it is risky, if a poor innovation is chosen). It is a mistake to presume that works best in a struggling school is the same as what works best in a successful school (Bendickson, et al., 2011).
8. Impact control. This is about where the locus of power is on determining the nature of outcomes from schooling. It has moved from developing the whole child, is now stuck in the literacy and numeracy phase, and soon will move to “we have a crisis as not every kid is above average” phase. The rhetoric, however, is 21st century skills (although we are 1/8th of the way through this century). The choices about how our schooling system impacts on students are a form of control increasingly being owned by the central agencies. As a consequence it is often narrow, given the imperatives for reliability and efficiency usually multiple choice tests finished in a few hours on one day a year (with the results coming back many months later), and become high stakes but with little improvement information for and about teachers or for students. (There are alternatives like the aSTtle system in NZ which is oriented to helping teachers understand their impact.)

These mechanisms can vary reasonably independently. Considering the recent changes, I suggest that the forces in Australian education have shifted over the past 40 years:

The pendulum swings in educational leadership



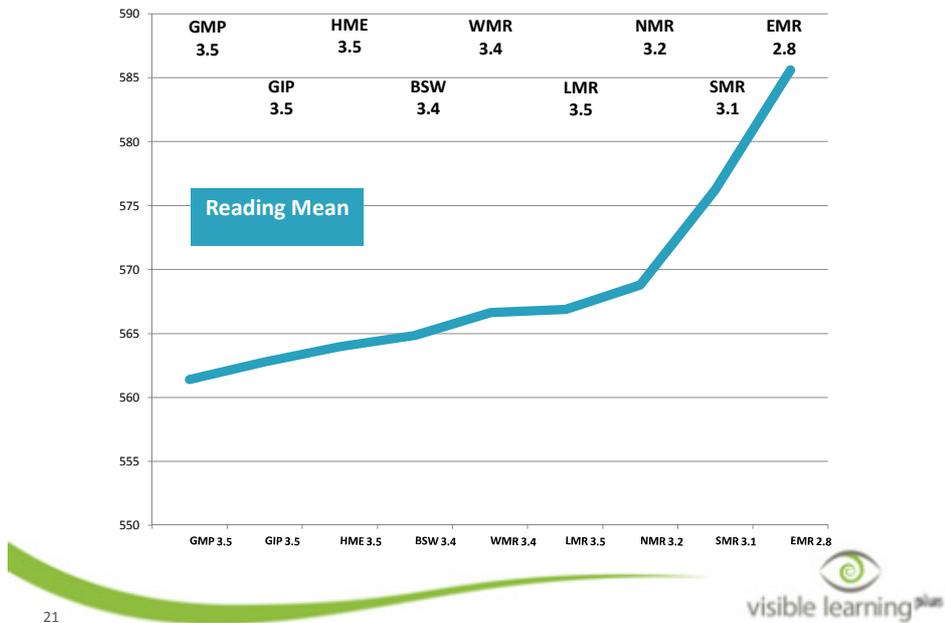
The major point of the above discussion is to ensure that autonomy is not seen as singular but as a polymorphous notion.

1. **Political control.** A common finding in the US Gallup polls is that 80% of parents consider the government education system as becoming worse, while 80% of these same parents consider the quality of education their children are getting in public schools is very good. The moves across Australia to independent government schools, increasing funding to non-government schools, and basing NAPLAN on average scores (which this best reflect the cost of housing in a school zone) promotes the false message that private schools are better and public schools are for the rest. This is not helped when school principals make claims about their local school that seems to parents as "we are better than them" and them is often the government school down the road. Coke and Pepsi do not advertise how bad the sickly sugary drink of the other is, but differentiate on quality of taste! How to introduce autonomy to local schools while have a consistent message about the quality of the overall school system is critical to enhance public perceptions about the quality of government schools, and to justify further investment in public schooling.
2. **Market control.** There is no inspector system in Australian schools, so the market is primarily informed by test score related information; although MySchool website is beginning to include many more sources of school data – in the spirit of transparency. There is not much evidence that parents choose schools based on this information; although many seek comfort in their local school by looking at this evidence (but rarely do they then move schools). As more autonomy is given to local schools (this is a political imperative) then the market needs more assurance that the local decisions are indeed enhancing the quality of schooling. The request is that two questions are asked of each local school:
 - What is the evidence that your school is having a substantial impact on the growth of all its students?
 - What is it doing in light of this evidence?

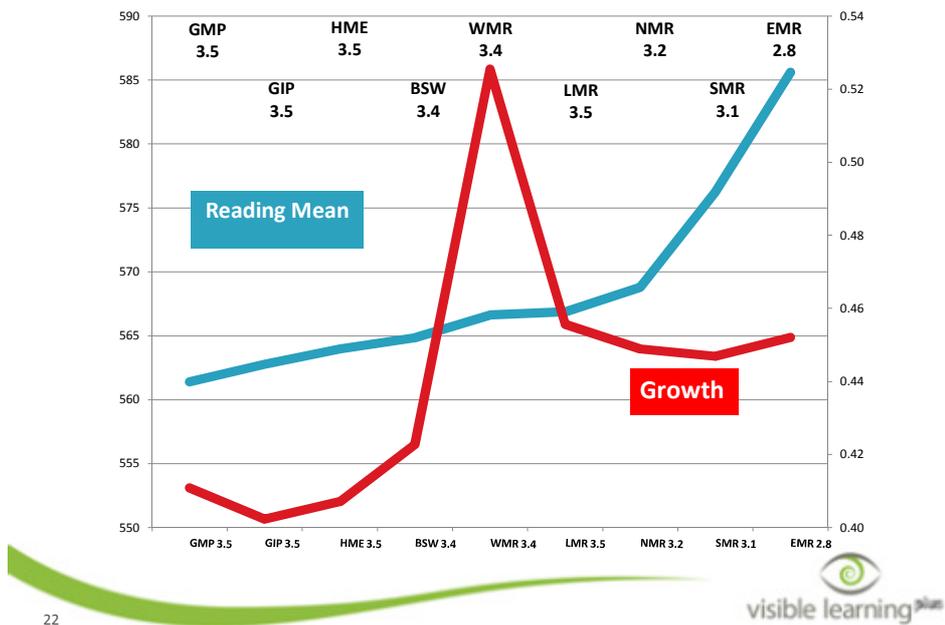
These two questions put local emphasis on the quality of evidence, and on evidence about actions in light of impact. They are premised on the notion that every students derives at least a year's growth for a year's input, begs to have more evidence that just test scores (even though they should be included), and the second question ensures that the focus is on improvement. This moves the debate from the current one of accountability about the past to accountability about improvement; allows parents and the community to be convinced by the range and quality of evidence; moves from a narrowing to just the easy to measure. But it should also require overview by education experts to ensure validity of the claims, validity of the interpretations, comprehensiveness of the evidence, and evaluation of the improvement.

3. Bureaucratic control. There are many reports of the high levels of administration and compliance now asked of principals. Some systems (e.g., US) have professional registrars in government schools that relieve much of this burden, and any move to provide more “autonomy” to schools may need a work-time analysis to see if this is the most efficient distribution of work. There may be many tasks better devolved to agencies that work across many schools although it is noted that when systems have trialled such devolving to other than principals there is much dispute. Principals tend to prefer to be project managers rather than devolve many of these functions – despite their protestations.
4. Social control. Is it government’s role to promote and fund a professional body of school leaders and professionals? There are such bodies in many other professions (Royal College of Surgeons, Board of Architects, Academies, etc.) but education has a multiplicity of disparate groups from Unions to local curricula groups (all serving legitimate and worthwhile purposes). But there is no group to recommend, consult, assist in implement, and to test and promote ideas. I note that in Ontario that government devolved all professional development budgets to a collective principals association and this has been a key part of their major improvements during the past 10 years. The principals are now accountable for the impact of PD and there is much more research by them as to quality, impact, and sustainability – they have moved from the traditional model of providing so many hours per school to being very adept at quality and quantity control. It is hard to justify devolving social control to 2500 schools in Victoria – the extra cost of principal time, the replication across so many schools, and the potential for mixed and contrary messages about government schools are too high. A key question in the autonomy debate is how to build a collective profession to assist in promoting the values and impact of government schooling, and to attain the efficiencies and effectiveness of the investment.
5. Curricula and Program control. It is a fundamental right in a democracy to make decisions about curricula and it is hard to imagine how this could be accomplished locally. As the US has shown when it had local control over curricula that debates about text books, intelligent design, and the place of religion often dominated debates about what was taught or not. There is a tipping point above which the quality of curricula hardly matters and Australia is well over this tipping point. The recent moves to a national curricula has again been resolved by “group think” as it is hard to find any evidence that the curricula was based on how students actually progress; and hard to find evidence about international benchmarking to the appropriateness of challenges in the curricula. (Note, if Australia is to become in the top 5 in the PISA world, it needs a curricula that enables our students to gain an extra year in reading and two years in numeracy by about age 10 – and this is unlikely given the current curricula). As significant is the hierarchy of curricula, particularly in high schools. There is still a major funnelling message about the subject chauvinism favouring university preparation and selection. Australia’s retention rate hovers around 80% and thus 1 in 5 students do not see staying in schooling worthwhile. To increase the retention rate does not mean narrowing and funnelling even more, but allowing schools to offer worthwhile programs where students can become excellent in coaching, service and trade industries along with their university bound peers. This would require a major change to the assessment structure to avoid the privileged and others, and I note NZ has made this assessment change with major increases (+10%) in retention rates over the past 8 years.
6. Pedagogical control. Many educators define their sense of professionalism in their autonomy to choose how to teach. There is, however, a “practice of teaching” as we know from many meta-analyses of research that some teaching methods are much more likely than others to have major positive effects on students – but say teachers, my class or my students are different (although the next teachers of these students make the same claim about the very same students that were ‘different’ in another teachers class). One of the reasons for the persistence of this claim for autonomy over teaching is that the standard of evidence is low – if a teacher can provide evidence that the students are learning we tend to accept the autonomy. But this is such a low standard, and as shown in Visible Learning, this standard is meeting by virtually every teacher. If the standard is set at the average of all influences on learning than the story is quite different. No teacher or school has the right to autonomy to teach leading to low impact (even if the impact positively but not sufficiently impact on learning). This is not suggesting there is one way to teach, but is saying that schools should demonstrate that all students make at least a year’s growth for a year’s input. This will lead to the discussion about accountability in the next section related to these notions.

7. Impact control. As the Victorian government edict is for more “autonomy” to principals then impact control becomes more important. Such control van not merely be the number of students who attain above some minimum standard as that will greatly advantage schools in areas where they already come to school at or above this standard. On the other hand, growth or progress is insufficient as some students need to make greater than average growth if they are ever to have the chance to share the spoils of school success. For example, using NAPLAN in reading, the following figure shows the *mean* reading score for each region – and this closely follows the socio-economic status of each region.

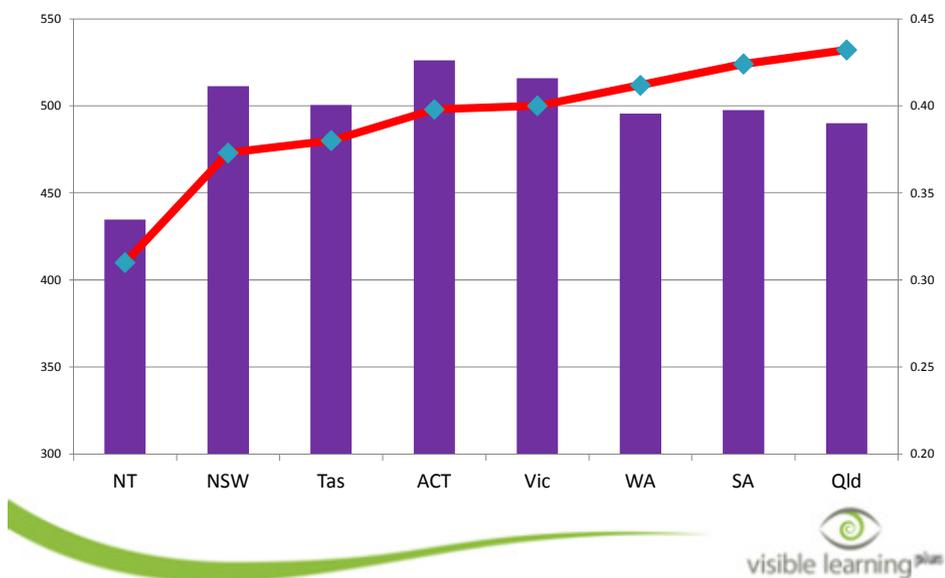


The following graph shows the average effect-size gain from 2008-2010 for each region and the message is quite different. Some of the more affluence regions are not adding much value whereas Western Region (in the middle above) is dramatically punching above expectations in adding value to their students.



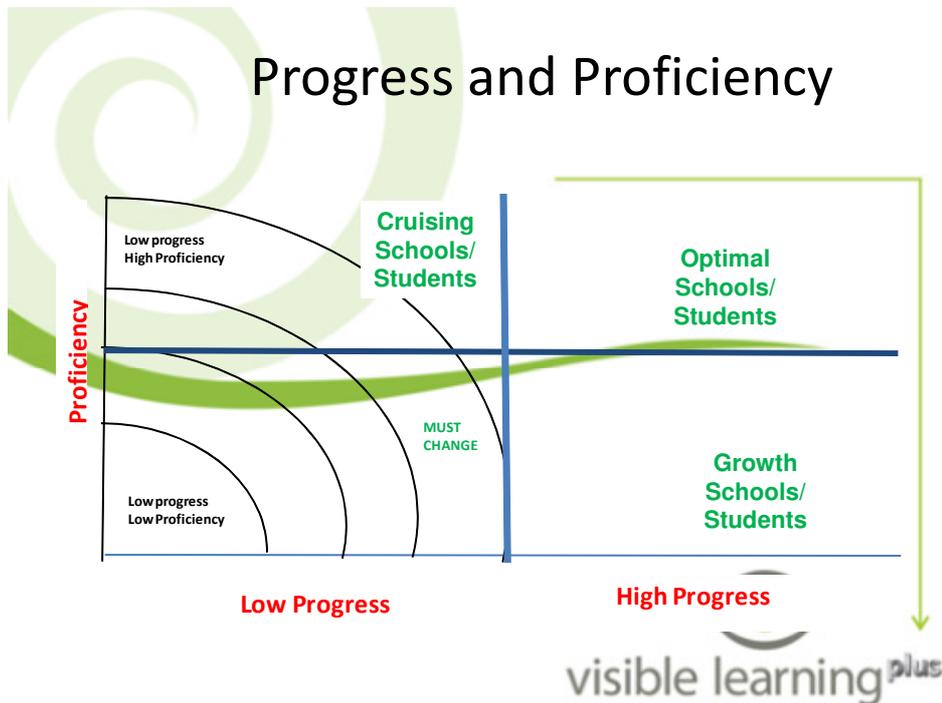
The similar state comparison shows Queensland ranks 7th in average NAPLAN but is far ahead of most states in value added. Both progress and proficiency are needed in any accountability model.

Average (purple) and Growth (red)



One possible model could be the following – schools could be asked to provide information about both progress and proficiency. They could be asked for this across many curricula domains (e.g., using NAPLAN standards for Proficiency, and effect-sizes > .40 as the average for progress). When I use this model using Victoria Secondary schools Yr 9 Reading, I get 27% in cruising, 10% in Growth, 45% in Optimal, and 18% in Must Change schools. These seems little defence for allowing the 18% of schools autonomy over much – they must

change and experience elsewhere shows they are hardly likely to change themselves. Autonomy must be earned.



Concluding comment

The above recognises the many meanings of autonomy in education. To declare that “A key direction of government policy in Victorian Government schools over the past two decades has been increased devolution or decision making to the school level” is thus misinforming at best and misleading at worst. Autonomy about what? And what are the consequences of changing the locus of decision making – relative to other decisions made elsewhere. If, for example schools are permitted control over budgets is it unfettered control (how much to pay the gardener, to spend more on football coaching than math and English) or within a series of policy guidelines. Are schools permitted control over curricula implementation but the centre retains control over assessment? Maybe a model of checks and balances is needed within which autonomy is discussed.

The evidence relating to autonomy

The road to autonomy is well travelled. There are many countries where, compared to the average, decisions related to resource allocation and curriculum and assessment are local (Czechoslovakia, Netherlands, Macao); where decisions relate only to resource and not curricula and assessment (Chile, Hungary, Sweden, Bulgaria, Dubai, and Shanghai); and where decisions are high relating to curricula and assessment (Japan, Korea, NZ, Hong Kong, China, Thailand).

The most remarkable outcome of high levels of increased devolution to schools is greater inequity. Given the premise of the model – the rich get richer and the poor get poorer (or close!) then it is inevitable. As good and bad decisions are made that then impact student learning this is an inevitable conclusion. Most systems that have increased devolution have led to more unequal schools. For New Zealand devolved much to schools as a consequence of the 1989 Tomorrow’s School mandates and since that time NZ has dramatically increased the inequalities in student outcomes – it is now known as “high quality, low equity”.

To be clear, low equity means that there is larger spread – the normal curve of achievement outcomes become flatter, the gap between the top and bottom is greater. It does not mean there are more students in “the tail” or that there are wider gaps that need to be closed (Hattie, 2010). Since 2000, the NZ system has taken back

aspects of devolution (e.g., one line budgets to schools) to begin to redress the inequalities. Another consequence is that the community start to believe proxies of school success more often to justify their decisions to send students to the schools, and crude indicators of socio-economic status get used as measures of school success (e.g., in NZ the 'decile system'), and schools present positive images of their success even when the evidence is overwhelmingly negative (e.g., South Auckland, East Coast, West Coast schools that need an external review to show the community contrary information – which was initially highly resisted). Even when some of the toxic schools were forced to close the community believed still that local decision making was a better alternative!

There is more detailed evidence relating to various aspects of autonomy. For example, the three typical decisions given locally are autonomy over personnel, within-school budgets, and hiring.

- Achievement is higher in countries where most schools have autonomy in staffing decisions and in hiring teachers, but negative when schools have autonomy over formulating their own budget – but the positive effects of the first two (staffing and hiring) only accrue when there are external exit examinations from the schooling system (Wobman, 2005).
- Achievement is higher when school principals have some control over opportunistic learning costs, such as purchase of instructional supplies, hiring and rewarding teachers (within a given budget), and choosing textbooks and instructional methods (Wobmann, 2001; Fuchs & Wobmann, 2007).
- Achievement was lower when teachers had some control over the amount of opportunistic learning costs but higher when they had some control over the resources to spend within the prescribed budget on (WoBmann, 2003).

These findings point to the obvious critical concern about whether to spend the resources on this rather than that – it is the efficiency of this distribution that is more critical than the amount. There is little evidence to suggest that the best performing systems have the higher amount of resources; and the current debates for “more” tend to miss this well established finding. Thus, there can be autonomy to make decisions about resources for instructional purposes as these tend to relate to achievement and learning outcomes, but there is little evidence that autonomy to make decisions about non-instructional purposes have little to no effect. The debate is what is instructional (text books, curricula materials) and what is not (class sizes, size of swimming pool).

Other related implications are suggested:

- The local autonomy is more effective when school leaders are instructional leaders (rather than transformational leaders (see Dinham, 2012; Hattie & Clinton, 2011)).
- When all students are “owned” by all teachers, there are high degrees of teacher collaboration about teaching and learning across the school, and there are high levels of discussion about teaching impact. Hoyle (1980) argued that achievement is increased where students are ‘owned’ by all teachers, and then teachers are willing to negotiate and exchange interests in the benefits of improving impact on students. He contrasted with “restricted professionals” who focus primarily on pedagogical concerns, subject matter, and their own teaching activities in their own classroom.

Conclusions

It is the argument in this paper that autonomy, in whatever form and however it is distributed, entails responsibilities. Autonomy is a polymorphous concept and thus should not be used as if it means one attribute. Providing schools with autonomy is misleading unless it specifies autonomy about what, with what responsibilities. Devolving *too many* responsibilities to schools (e.g., over curricula, personnel, and budgets) has lead to increased inequalities within the schooling system; distortion of information about impact; and the masking of failing schools.

1. Autonomy can relate to political control, market control, bureaucratic control, social control, curricula and program control, pedagogical control, and impact control. Decisions about autonomy require bringing these various forms of control into equilibrium to maximise impact on learning.
2. Ensuring all schools are working to create a common and powerful narrative about the impact of schools on students such that it has high convincability and credibility not only to themselves professionally but

also to their various communities – using quality evidence about impact on increasing proficiency and progress. Assuming that schools have expertise in market design about their impact is false; although they have shown high levels of expertise in marketing an image of their impact.

3. With guidance, increased autonomy can be coupled with responsibilities to provide quality evidence about impact and learning from this evidence. The rhetoric is more powerful about system improvement build on system success.
4. A careful review of bureaucratic and compliance costs should be undertaken as part of any devolution to schools, and questions asked about “who” in schools are best suited to enacting these bureaucratic and compliance matters (and it is unlikely to be the highest paid person in the school, the principal).
5. The Victorian government should consider its role in developing, promoting, and sustaining a professional body of teachers and school leaders to be a key part in improving Victorian schools, critiquing and successfully implementing policies, and being responsibilities for improved schooling.
6. Schools should have autonomy to find ways to make their schools inviting places for all students to attend, and stay till the final year - and provide evidence of this success of their curricula and program implementation. But it may require a rethink about the current assessment constraints that privilege tertiary bound students and teachers and parents who aim for tertiary programs.
7. Schools should have autonomy over the quality and nature of evidence of the impact of their teaching decisions – that is shared across schools, relates to progress and proficiency, and leads to consequential decisions to retain or change teaching programs.
8. The system should retain control over the forms of evidence needed to hold schools accountability, schools should be given greater autonomy over choosing the nature of evidence to meet this accountability, and there needs to be central controls over the quality of this evidence and over the worthwhileness of the interpretations by schools to enhance or change their programs.
9. Such evidence should relate both to proficiency and progress for ALL students.
10. Schools should have more control over budgets specifically related to learning expenditures

The unintended consequences of high levels of devolution or autonomy are multiple and it seems remarkable that governments continue this strategy of promising “autonomy” when the critical issue is autonomy about what, within what constraints and responsibilities, and to whom. School governance is about equilibrium; and it is barely conceivable that Victorian public schools will make the improvements required by letting 2500 flowers bloom.