

Synlig læring i dag



John Hattie

Professor og leder af Melbourne Education Research Institute på University of Melbourne i Australien og vicedirektør for Science of Learning Research Centre¹

I bogen *Visible Learning* fra 2009 skrev John Hattie læringshistorie i form af verdens hidtil mest omfattende forskningsoversigt: en syntese af over 800 metaanalyser af faktorer, der påvirker elevernes læring. I denne artikel fortæller han historien om synlig læring og forholder sig til en række af de kritikpunkter, der siden er blevet fremsat.

Synlig læring omfatter på nuværende tidspunkt en serie bøger og publikationer, der sigter på at vise de budskaber fra forskningen, som kan præge praksis i undervisning og læring. Tilskyndelsen til dette arbejde begyndte i mine tidlige universitetsår (de sene 1970'ere) af to grunde. For det første handler min ph.d. om målemetoder (Hattie 1984, 1985). Jeg opdagede hurtigt, at en psykometriker var velkommen i læreruddannelsen, men jeg fik også at vide – på en måde der ikke var til at tage fejl af – at det kun kunne lade sig gøre, hvis jeg også var i besiddelse af en faglig kompetence inden for et væsentligt pædagogisk domæne. Alle mine kolleger bad mig indtrængende om at studere deres yndlingsområde, fordi deres yndlingsområde uden for al tvivl var netop det, der gjorde en forskel for elevernes læring (men de var alle forskellige!). Umiddelbart efter min ph.d. førte mit akademiske job mig til en involvering i læreruddannelse. Det budskab, jeg hørte fra denne sektor, var klart: Ignorer alt, hvad du lærer i elfenbenstårnet,

og "se så bare, hvad jeg gør" (eller varianter af dette). Men hvordan kan det være, at alle lærerne var i stand til at frembringe "evidens" for, at deres metode havde en positiv effekt på elevernes læring, når mange af dem var så forskellige og i virkeligheden også modstridende? Det er ikke så mærkeligt, at de lærerstuderende frit kunne vælge den metode, der appellerede mest til dem, uanset hvilken "forskningsevidens" der forelå.

For det andet havde Gene Glass (1976) netop publiceret metaanalysens metodologi, og det var spændende. Den bedste måde at forstå denne metode på var at gennemføre en metaanalyse, så Brian Hansford og jeg gik i gang med et studie af relationen mellem selvopfattelse og præstationer (Hansford og Hattie 1982a, 1982b). Dette førte til en langvarig interesse for metaanalyser, og jeg fortsatte med at producere dem og indsamle dem. Det slog mig med forbavselse at opleve deres simple og magtfulde statistiske styrke, men også at se, at deres resultater ikke gjorde nogen større forskel på undervisningsområdet. Dette skyldes primært, at mange af dem var rige på data, men fattige på mening. Resultaterne var ikke indlejret i en mere substantiel litteratur, og konklusionerne syntes at blive hængende i bestemmelser af, om den generelle effekt var lav, middel eller høj – og selv den mindste effekt blev anset af mange som "høj" eller i det mindste tilstrækkelig højt over nul til at være af interesse og give mening. Endvidere blev det tydeligt, at så godt som alle metaanalyser viste positive effekter.

På den måde startede en hobby – at indsamle metaanalyser og prøve at udarbejde en fælles historie, der kunne binde de mange resultater sammen. Selvom

¹ Artiklen er oversat af Søren Søgaard.

jeg ikke fik gjort historien tilstrækkeligt overbevisende i disse tidlige år, har de vigtigste temaer i dette arbejde overlevet (Hattie 1987, 1992). Det første tema er den "relative" betydning af effekter af påvirkningsfaktorer på læring. Dette betyder, at vi må spørge efter den relative virkning og ikke sammenligne med fraværet af effekten. Alt for ofte sammenligner vi fejlagtigt en påvirkning (for eksempel et undervisningsprogram) med fraværet af denne påvirkning. Dette er at sammenligne med nulpunktet, mens vi burde sammenligne denne påvirkning med alle andre mulige påvirkninger. Denne "relative" effekt kan udregnes ud fra akkumuleringen af alle mulige påvirkninger på elevpræstationer – hvilket fra Visible Learning-arbejdet er $d = 0,40$. Så enhver effekt, der er større end 0,40, er over gennemsnittet, mens mindre end 0,40 er under gennemsnittet.

Det andet tema er behovet for at udvikle en historie om den relative virkning af disse påvirkninger over og under dette gennemsnit. At udvikle denne historie er det, der tog 20 år og førte til bogen *Visible Learning* i 2009. Jeg blev derefter bedt om at skrive en mere tilgængelig bog for lærere (*Visible Learning for Teachers* 2012; dansk udgave: *Synlig læring – for lærere* 2013). Mange bad derefter om flere detaljer om hver af påvirkningsfaktorerne, hvilket førte til mit samarbejde med Eric Anderman om at redigere *The International Guide to Student Achievement* (Hattie og Anderman 2013). Samtidig ønskede jeg at dykke dybere ned i "lærings"-aspekterne i *Visible Learning*. Greg Yates og jeg publicerede derefter en bog om synlig læring og videnskaben om, hvordan vi lærer (Hattie og Yates 2014; dansk udgave: *Synlig læring og læringens anatomi* 2014).

Det tredje tema hænger sammen med, at jeg var blevet bedt om at præsentere arbejdet for mange forskellige typer publikum (lærere, ledere, planlæggere, politikere). Det gik hurtigt op for mig, at jeg kunne

tale med nogen autoritet om forskningsresultaterne, men implikationerne for praksis var ikke min stærke side. På det tidspunkt var jeg leder af udviklingen af New Zealands nationale vurderingssystem og havde et fremragende team af forskere og praktikere til rådighed (se Hattie og Brown 2008). Mit ønske til dem var at hjælpe mig med at føre historien om synlig læring ud i praksis, og under ledelse af Debra Masters har dette team præsteret dette i en sådan grad, at budskaberne er gået verden rundt. Mange skoler har implementeret budskaberne, og vores næste bog sætter fokus på 15 casestudier af skoler, der succesfuldt har implementeret denne historie (se herunder; Hattie, Masters og Birch 2015).

Jagten på moderatører

Igennem historien om min og mange andres undersøgelser af de mest virkningsfulde påvirkninger af elevpræstationer, har der været en fælles mission – og det er at opdage moderatører, hvilket er de faktorer, som modificerer de generelle gennemsnitseffekter (jævnfør Cronbach og Snow 1977). For eksempel: Gør alder en forskel på effekterne af klassestørrelser (er effekterne i indskoling og på mellemtrin større end på overbygningen)? Gør typen af vurderingsresultater en forskel (overflade versus dybere resultater)? Gør påvirkningen fra udviklede lande og udviklingslande en forskel og så videre. Jagten på moderatører opslugte mig under udviklingen af synlig læring-historien. Det, der virkelig overraskede mig, var, at der var så få af dem i forskningslitteraturen, men her er det også meget vigtigt at bemærke, at meget af forskningen var langt mindre opmærksom på moderatører, end man skulle forvente. De moderatører, der var lettest at måle (køn, alder, socioøkonomisk status) var til stede, men de mere afgørende moderatører (kulturelle effekter, tidligere præstationer, tilbøjelighed for at lære) var sjældne. Hvor moderatører rent faktisk var til stede, blev de pointeret (for eksempel de forskellige effekter af lektier i

grundskolen og i gymnasiet). Denne eftersøgning må fortsættes, og kvaliteten af moderatorerne i den oprindelige forskning og i metaanalyserne kunne blive langt større.

På samme måde er variabiliteten af gennemsnitene fra de mange metaanalyser vigtig. At udregne den variation, der er forbundet med hver af påvirkningerne, er et vigtigt anliggende, når man gennemfører en metaanalyse, og metoderne inkluderer en evaluering af graden af heterogenitet hen over studierne og en vurdering af, om gennemsnittet er en rimelig typisk måling (se Borenstein, Hedges Higgins og Rothstein 2011; Huedo-Medina, Sanchez-Meca, Marin-Martinez og Botella 2006). I *Visible Learning* kommenteres de variationer, der påvises for de enkelte påvirkninger, og hvor de er særligt store. For eksempel er variationen i effekterne af feedback særligt store, og derfor kan enhver generel konklusion, som er baseret på gennemsnittet, være vildledende. Jeg har brugt megen forskningstid på at prøve at forstå de stærke effekter af feedback, og simple svar er ikke tilstrækkelige (Gan og Hattie 2014; Harks, Rakoczy, Hattie, Besser og Klieme 2013; Hattie 2013; Hattie og Timperley 2007). For eksempel er det af afgørende betydning at sondre mellem at give og at modtage feedback, mellem feedback, der handler om, "hvordan klarer jeg mig", og "hvor skal vi hen herfra", og feedbackeffekterne fra eleverne til lærerne.

Betydningen af referencepunktet $d = 0,40$

Referencepunktet 0,40 er simpelthen gennemsnittet af alle tænkelige påvirkninger – uden at være tyngt af nogen moderator-effekter. Så enkelt er det. Nogle virker overraskede over, at effekten har moderatorer – for eksempel er det velkendt, at effekterne af læse- og skrivefærdighed og talfærdigheder er større i grundskoler end i gymnasier; og at effekterne kan være større, når resultaterne er udtryk for en overfladeforståelse mere end en dybere forståelse. Dette

gør ikke gennemsnittet ugyldigt. Det viser blot, at skoleniveau og resultaternes beskaffenhed kan optræde som moderatorer. Dette betyder imidlertid ikke, at gennemsnittet på 0,40 eller derover ikke kan opnås ved implementeringer i gymnasiet.

Det betyder heller ikke, at blot det at placere en lærer foran en klasse vil føre til en forbedring på 0,40. Bevidste forsøg på at forandre, forbedre, planlægge, modificere eller innovere er involveret.

Jeg har brugt uafhængige estimater til at spørge om den gennemsnitlige udvikling af elever i skoler over et år. I Australien er der for eksempel nationale tests i læsning, skrivning og talfærdigheder (NAPLAN), og ud fra en analyse af vækstdata (samme elever målt hvert andet år) er den gennemsnitlige årlige effektstørrelse 0,40 – den gennemsnitlige elev har således en samlet fremgang på 0,40 per år. Effekten er højere i grundskolerne (0,55) og lavere i gymnasieskolerne (0,25). Så man skal være forsigtig med at overgeneralisere 0,40-effekten – men den tjener dog som et nyttigt referencepunkt med hensyn til både at forstå faktorerne over og under dette referencepunkt; og den giver os i hvert fald en begyndelse på samtalen om, hvorvidt eleverne modtager mindst et års vækst for et års input.

Historien om synlig læring: Dens betydning i klasseværelserne

Det vigtigste budskab er: "Kend din virkning". Det vil sige: Indtag en professionel evaluators holdning, og spørg efter interventionernes fordele, værdi og vigtighed – i klasseværelset, blandt personalet og i den politiske sektor. Undervisernes rolle er ikke blot at indsamle data, skrive rapporter og udfylde spørgeskemaundersøgelser, men at være fremragende *fortolkere* af evidens. Dette betyder, at der må være en kultur for at søge evidens til støtte for fortolkninger om virkninger og om betydningen og dybden af "virk-

ninger" på eleverne (af lærerne, og af de førte politikker). En sådan evaluering kræver, at underviserne foretager bedømmelser af deres virkning og deres lederskab sammen med andre i denne evaluering af virkninger, såvel som bedømmelser af deres praksis. Den grundlæggende præmis i modellen for synlig læring er, at når underviserne fokuserer på at definere, evaluere og forstå deres virkning, fører det til en maksimering af elevernes læring og præstationer.

De vigtigste præmisser for synlig læring inkluderer, at lærerne tror på, at deres vigtigste rolle er at evaluere deres virkning, samarbejde for at kende og evaluere deres virkning, klart kende elevernes tidligere læring, være eksplicite over for eleverne ved begyndelsen af et undervisningsforløb med hensyn til, hvad succes i dette forløb vil sige, implementere programmer, der har den optimale fordeling af overfladelæring og dyb læring samt have passende niveauer af udfordringer og aldrig acceptere, at eleverne blot "gør deres bedste". Dette betyder, at lærerne må anlægge en evalueringsholdning: fremme forbedringer, evaluere virkninger og gøre evalueringer af eller reaktioner på interventioner til deres vigtigste tankesæt. Dette involverer at være tydelige om deres kriterier for succes, om den passende balance mellem overfladeviden og dyb viden som en del af disse succeskriterier, at forstå, hvor eleverne befinder sig i forhold til målene, samt at sikre, at succeskriterierne er tilstrækkeligt udfordrende for alle elever. Et centralt spørgsmål er, hvordan man skal opbygge lærernes og skoleadministratorernes kapacitet til kollektivt at udvikle og evaluere succesfulde undervisningsprogrammer og læringsoplevelser (for en mere detaljeret forklaring, se Clinton og Hattie 2014).

Det er således en disposition til at ønske at kende sin virkning, der ligger i kernen af VL¹-modellen. Det er

ikke at bruge interventioner med høje snarere end lave forventningsrater, krydse toppen af og ignorere bunden – det er i stedet et tankesæt, der siger: "Når jeg går ind i et klasseværelse, er det min rolle at evaluere min virkning" – og eleverne nyder godt af det. Den største effekt har at gøre med lærernes ekspertise i at evaluere deres virkning (læg mærke til flertalsformen, fordi det som oftest kræver en gruppe af lærere under ledelse af skolelederen at evaluere virkningerne, dybden af virkningerne og sammenligneligheden af virkningerne på tværs af hele skolen og mange gange på tværs af flere skoler).

Denne betoning af at kende sin virkning dikterer ikke specielle metoder, arbejdsprogrammer eller klassestrukturer. Den fremhæver i stedet undervisernes kapacitet til at designe effektive programmer, implementere dem med kvalitet og derefter kritisk bestemme størrelsen af deres undervisningsprogrammers virkning på elevernes læring. Denne betoning af forandringens størrelse er af afgørende betydning, især i betragtning af, at mere end 95 procent af de metoder, der anvendes for at fremme læring, fører til positive virkninger på læringen (Hattie 2009). Blot at fremme læring er ikke nok – for at opnå maksimale resultater for eleverne må læreren kende størrelsen af sin virkning og derefter evaluere, om denne virkning er tilstrækkelig.

Den vigtigste historie kan opsummeres i seks vigtige forskningsresultater. Læg mærke til, at der ikke findes en specifik metaanalyse, der fører til disse estimater, fordi de er baseret på gennemsnitsberegninger på tværs af mange metaanalyser. For det første forstærkes virkningen, når lærerne forstår, at deres vigtigste rolle er at evaluere deres virkning ($d = 0,91$), når lærerne arbejder sammen om at kende og evaluere deres virkning ($d = 0,91$), når lærerne klart kender deres elevs forudgående læring (hvad de bringer med sig til timen, $d = 0,85$), når lærerne ved begyndelsen af et

1 VL = Visible Learning. O.a.

undervisningsforløb eksplicit informerer eleverne om, hvad succes består i ($d = 0,77$), når lærerne implementerer programmer, som har den optimale fordeling af overfladelæring og dyb læring ($d = 0,71$), og når lærerne anlægger passende udfordringsniveauer og aldrig forventer, at det er nok at "gøre sit bedste" ($d = 0,57$).

En sådan virkning måles ikke kun i ændringer af testscorer, selvom testscorer kan fungere som feedback til lærerne om deres virkning. Men også af de niveauer af investering, mestring, tolerance over for fejl, grader af samarbejde med andre for at lære mere, iver efter at få succes og fremkaldelse af en høj interesse hos eleverne – hvilket er mere tilbøjeligt til at følge af succes med læring og ikke nødvendigvis er en forudsætning for at engagere sig i læring. Et højere engagement er mere et resultat af læring, end det er et nødvendigt input.

Denne model beder underviserne om at evaluere kvaliteten af den evidens, de kan tilvejebringe for fire vigtige temaer: 1) strategisk planlægning og selvkritik, der inkorporerer feedback om interventioner og virkning af undervisningen; 2) tilstedeværelsen af bevidste strategier for styrkelse af undervisningens virkning, som overvåger det evaluerende tankesæt hos lærerne; 3) brugen af elevernes stemme som en del af responsen på interventionerne (det vil sige om virkningen); og 4) indsamling og analyser af data om virkningerne. Det akronym, lærerne bruger, er DIE: Det vil sige at de deltager i en Diagnosticering af elevernes status, når de påbegynder lektionerne, at de har mange forskellige Interventioner, de kan anvende, hvis deres igangværende intervention ikke har den ønskede virkning, og at de Evaluerer elevernes respons på deres interventioner.

Der er andre meget vigtige tankesæt, men de følger alle af dette første "Kend din virkning". De andre in-

kluderer: Jeg er en forandringsagent; jeg informerer fra begyndelsen eksplicit eleverne om, hvad en succesfuld virkning er; jeg ser vurdering som en tilvejebringelse af feedback om min virkning; jeg arbejder sammen med andre lærere om at udvikle fælles opfattelser af fremskridt; jeg engagerer mig i dialoger, ikke monologer; jeg stræber efter udfordringer og ikke efter at "gøre mit bedste"; jeg bruger læringens sprog; og jeg ser fejl som muligheder for læring. Bogen *Visible Learning into Action* (Hattie, Masters og Birch 2015) beskriver implementeringsmodellen for opnåelse af den størst mulige virkning mere detaljeret, og Clinton og Hattie (2014) beskriver synlig lærings 10 principper som en myndiggørende implementering af evaluering som redskab.

Ideen om undervisere som evaluører "implicerer bevidste forandringer, styring af læringen og at gøre en synlig forskel for elevernes oplevelser og resultater (og for lærernes) – og nøglemekanismen for denne aktivering er et tankesæt, der favner rollen som evaluører" (Hattie og Clinton 2011:99). Det er således meget vanskeligt at forestille sig, at den enkelte lærer kan implementere synlig læring som en isoleret ø; det kræver i stedet høj grad af intervention på skole- og netværksniveau. Skoleledere har brug for at legitimere debatten om virkningerne i deres skole, skabe et trygt og fair miljø, hvor dette kan finde sted, og levere ressourcerne (for eksempel tid) til kollektivt samarbejde i denne jagt på den højest mulige virkning på alle elever. Som Senge (1990) bemærkede, kræver dette skoleledere, som er oprigtigt interesseret i praksisens verden, som yder denne verden den højeste respekt og er oprigtigt interesseret i at bruge evaluerende metoder til at gøre skolen bedre for alle.

"Det er ikke længere tilstrækkeligt at have en enkelt person, der lærer for hele organisationen, en Ford eller en Sloan eller en Watson. Det er ikke længere muligt at 'finde ud

af det' fra toppen og få alle andre til at adlyde ordrerne fra den 'store strateg'. De organisationer, der virkelig vil præstere fremragende i fremtiden, vil være de organisationer, der udnytter menneskers engagement og kapacitet til at lære på alle niveauer i en organisation." (Senge 1990:4)

Endvidere hævder Hattie og Clinton (2011), at der er mange evalueringsroller for skolelederne: at skabe mening, opstille prioriteringer, sikre lighed, coaching, stille og rammesætte spørgsmål samt evaluere deres egen virkning på lærere og elever. Lederens rolle involverer også at bringe ekspertise til de forskellige interventioner, udvikle partnerskaber, engagere lokalsamfundet, udvikle evalueringskapacitet, formidle viden ind i skolen, fremme tænkning samt kontinuerligt at være kvalitetsbevidst.

Synlig lærings implikationer for forskningen

Jeg blev uddannet til at mene, at fremskridt inden for forskningen kommer fra et tankesæt, der søger falsificering. Denne idé stammer fra Popper (1968) og går ud på, at man fremsætter en teori (for eksempel VL-modellen herover) og derefter spørger, hvilken evidens der ville overbevise mig om, at jeg tager fejl (det vil sige søge modstridende evidens).

"Dristige ideer, ubegrundede forudsigelser og spekulativ tænkning er vores eneste måder at fortolke naturen på: vores eneste *organon*, vores eneste instrument til at begribe den. Og vi må driste os til at fremsætte dem for at vinde vores pris. De blandt os, som er uvillige til at eksponere deres ideer til risikoen for gen-drivelse, er ikke en del af den videnskabelige verden." (Popper 1968:280)

Der findes, selvfølgelig, mange andre fortolkninger af de (nu) mere end 1.200 metaanalyser, og jeg har

leveret den opsummering af data, som vil gøre det muligt for mange andre forskere at fremsætte alternative forklaringer. Og ja, VL-modellen er spekulativ, men vigtigst er det, at den antyder konsekvenser og handlinger, der kan testes i klasseværelserne (som beskrevet i Hattie, Masters og Birch 2015). Og det er kun en teori til forklaring af data og kan ganske rigtigt vises at være ukorrekt eller forbedres.

Ud fra læsningen af de mange metaanalyser og i visse tilfælde mange af de oprindelige studier er der meget i vores forskningsmetoder, der kan forbedres. Moderatorernes kvalitet er allerede blevet bemærket, og behovet for, at de moderatører, der er involveret i metaanalyserne, kan levere fortolkninger og ikke blot data, er afgørende for fremskridt i vores forståelse. Den seneste brug af metaanalyser til udforskning af konkurrerende forklaringer er spændende. For eksempel modstillede Rowland (2014) forskellige forklaringer på det virkningsfulde ved testningseffekten (ved at levere mange muligheder for at praktisere tests). Han beskrev de vigtigste teorier, der forklarer denne effekt. *Genkaldelsesteoriene* antager, at en forstærket eksponering over for et materiale fører til større læring, bedre lagring og genkaldelse, og testningseffekten kommer af den indsats, intensitet eller processeringsdybde, der fremkaldes af de indledende tests. Hypotesen om den *elaborerede genkaldelse* hævder, at mange eksponeringer kan forøge antallet af genkaldelsesveje, der så kan udnyttes senere, fremmer speciel eller emnespecifik processering af testet information og gør det muligt at elaborere et specifikt stykke information. Hypotesen om *medieringseffektivitet* hævder, at gentaget testning kan styrke forbindelsen mellem stikord og den ønskede information. Hypotesen om *genkaldelseeffekten* antager, at testning bevirker en væsentlig forøgelse af de elementer, der kan genkaldes med succes. Endelig hævder teorien om *overførsel af relevant processing*, at forøgelsen stammer fra det overlap i pro-

cesseringen, der finder sted under den indledende og den afsluttende testning. De 61 studier i hans meta-analyse gav en gennemsnitlig effektstørrelse på 0,50 og viste en markant virkning af gentagne testninger. Mere interessant var det, at der var mest støtte til teorierne om *genkaldelseeffekten*, især når der var feedback under de gentagne testninger. Når der ikke var nogen feedback, var effektstørrelsen 0,03, nogen feedback gav 0,29, moderate niveauer af feedback 0,56 og megen feedback 0,73.

Den mest spændende udvikling er, at metaanalysernes felt er blevet bredere og nu inkluderer synteser af kvalitativ forskning og placerer metaanalyser som blot en af metoderne til at syntetisere forskning (Suri 2014). For eksempel gennemgik Kennedy (2008) forskellige kvalitative studier vedrørende relationen mellem lærer kvalifikationer og undervisningskvalitet og konkluderede, at denne "relation burde være selvindlysende – men overhovedet ikke er det" (s. 345). Hun brugte 20 kategorier, der relaterede til lærernes uddannelsesbaggrund, og 20 kategorier for evidens for undervisningens kvalitet. Af de lokaliserede 450 studier udelukkede hun dem, der ikke var baseret på virkelige klasseværelser (eksempelvis simuleringer), dem, der ikke etablerer en forbindelse mellem kvalifikationer og kvaliteten af praksis (eksempelvis kun beskrev begge og hævdede, at de var forbundne) – og dette førte til et endeligt korpus på 23 studier. En sparsom samling studier af et så vigtigt emne, som så mange har udtalt sig om. Kennedy konkluderede, "at den kvalitative litteratur er enig med den kvantitative litteratur om umuligheden af at skelne mellem lærere med forskellige certificeringer og forskellige læreruddannelsesbaggrunde" (s. 348). Hun gennemgik også de kvalitative studier af effekterne af læreruddannelse, og konklusionen var, at der ikke var større mærkbare effekter af disse programmer på lærernes praksis og resultater. Når lærerstudierende fra eksemplariske programmer blev sammenlignet med andre

studerende, var de første mere sikre på deres evne til at undervise, mere tilbøjelige til at omtale specifikke dele af deres programmer, som de satte pris på, og mere tilbøjelige til at basere undervisningsbeslutninger på elevernes behov snarere end på læreplanens krav. Men "ingen af studierne af læreruddannelserne fandt en klar, synlig virkning på undervisningspraksis" – og hun foreslår, at det muligvis skyldes, at de fleste programmer ønsker at ændre de ting, lærerne gør (for eksempel anlægge en bestemt tilgang), men at disse påvirkninger kræver færdigheder, som ligger ud over det, nybegyndere er i stand til at gøre (for eksempel spontane justeringer midt i en lektion).

Kritikken af synlig læring

Alexander (2010) bemærkede ændringer i den akademiske kritiks beskaffenhed over de sidste tiår. Han identificerede fire nedbrydende kilder til kritik: "dikotomiens, spottens, mytens og meningsløshedens diskurser". Dikotomiens diskurs reducerer alt til gensidigt udelukkende modsætninger, spottens diskurs definerer tonen i responsen ("først forvanske, derefter gøre nar, gøre angrebet personligt, hvor det er muligt, og appellere til den laveste fællesnævner i de folkelige fordomme"). Mytens diskurs, som mange kritikere benytter sig af, er, at "sådan var det ikke i min tid", "jeg underviste ikke på den måde", "vi benyttede os ikke af jargon eller managementsprog i vores tid", samt en generel amnesi over for vanskelighederne ved at implementere nogen positiv forandring. Meningsløshedens diskurs handler om ødelæggelse af selve sproget – "nye initiativer, trinvis forandringer, turbostart, løfte standarderne, innovere, nul tolerance, bedste praksis, skoler i verdensklasse, back to basics og alt det andet". Denne diskurs har altid været til stede.

Det skal imidlertid bemærkes, at kritik er essensen af den akademiske forskning og kan føre til en mere kvalificeret debat. Det grundlæggende ved den aka-

demiske kritik er, at den forholder sig til ideen, ikke personen; overvejer teser og antagelser omhyggeligt; og sigter på at informere og udvide debatten – ikke lukke den ned, fordi ideerne ikke er velkomne eller ikke er en del af visse kritikeres dogmer. Kritik bør ikke sætte falske præmisser eller forfægte detaljen frem for budskabet, og det kræver i det mindste et minimum af høflighed og god vilje at fremme debatten.

På trods af de overvældende positive kommentarer om synlig læring i de mange anmeldelser af bøgerne, har der også været kritik. Megen kritik er baseret på et eller andet aspekt af data, mens der er færre, som kritiserer historien. En del af kritikken handler om fejlfortolkninger af arbejdet, en del om udtalelser, jeg aldrig har hørt eller fremsat (eksempelvis det absurde tweet, der hævdede, at jeg skulle have sagt, at "halvdelen af data i *Visible Learning* er forkerte!"), nogle foretrækker at angribe mig personligt (hvilket ikke hører nogen steder hjemme inden for akademisk forskning), og nogle handler om mindre fejl (som jeg byder velkommen og retter i hvert optryk, hvor det er muligt). Nogle har kritiseret mig for enten at ignorere eller ikke at kunne se, at det er fattigdom, familier, hjemmemiljø og resurser, som er de vigtigste påvirkninger – der er to kapitler om disse påvirkninger i *Visible Learning* (Hattie 2009).

Der er begrænsninger for generaliseringerne. Evidensen er hovedsagelig baseret på 4- til 20-årige, og således er overførsler til voksnes læring problematisk – disciplinen mangler stærkt et lignende studie, baseret på voksne. Evidensen stammer hovedsageligt fra udviklede vestlige lande, og jeg bemærker i VL, at en generalisering til lande, hvor variationen i skolerne er meget stor – som den typisk er i mange udviklingslande – er problematisk. Igen understreges betydningen af moderatorer. Evidensen er baseret på elevpræstationer, og der er mange andre vigtige resultater af skolegang! Skoler handler så absolut om

meget andet end præstationer. Der er i øjeblikket et vigtigt forskningsprogram i Tyskland, som er i gang med en metasyntese af studier i motivation og affektive resultater, jeg er ved at gennemføre en metasyntese af studier med læringsstrategier som resultater, Mitchell (2014) har gennemført en metasyntese af studier, som er baseret på elever i specialundervisning (se også Swanson 1999), og jeg ser frem til en metasyntese af studier i klasseværelsesobservation.

I denne artikel vil jeg beskæftige mig med fire almindelige typer kritik.

1. Eksistensen af effektstørrelser

Eksistensen af effektstørrelser og metaanalyser er blevet draget i tvivl, på trods af at denne metodologi er vel etableret. Effektstørrelsesstatistikker er blevet sofistikerede, og der er mange bøger om dette emne (Cooper, Hedge og Valentine 2009; Hedges og Olkin 1985; Lipsey og Wilson 2001; Schulze 2004). Nogle kritikere har hævdedet, at matematikere ikke bruger effektstørrelser, at de ikke findes i matematiske lærebøger, og at statistiske samlinger ikke kalkulerer dem. Selvom effektstørrelser primært hører hjemme inden for disciplinen statistik snarere end matematik, er metoden almindelig i statistiske lærebøger, og de mest brugte statistiske samlinger kalkulerer dem faktisk (og der er mange internetsider, der viser, hvordan man udregner dem). Faktisk ligger begrebet "forskelle i standardafvigelser" i kernen af de fleste statistiske tests. Det er imidlertid vigtigt at bemærke, at der er mange måder at udtrykke en effektstørrelse på (eta-squared), mange kontroverser (som for eksempel om det optimale estimat af den samlede varians i nævneren) og mange analysemodeller (brugen af randomiserede eller faste effektmodeller, bayesiske metoder osv.). Effektstørrelser er grundlæggende standardmålestokke, og deres eksistens er veletableret.

Selvom visse mindre problemer er blevet anerkendt inden for VL (eksempelvis udregningen af CLE (Common Language Effect Size), og disse er, hvor det har været muligt, blevet rettet i det næste optryk eller oversættelse), er det ikke tilfældet og er aldrig blevet hævdet af nogen i VL-teamet, at "halvdelen af data" er forkerte. Dataene er baseret på en beregning af effektstørrelser ud fra, hvad der i øjeblikket er mere end 1.200 metaanalyser, og at sige, at halvdelen af dem er forkerte, ville være at beskyldte disse 1.200+ forfattere for at tage fejl. De opsummerede data har været offentligt tilgængelige fra begyndelsen, så alle kunne tjekke dem.

Endvidere er der to hovedtyper af effektstørrelser – dem, der er baseret på en sammenligning af grupper (som for eksempel en klasse, der modtager en bestemt behandling – som eksempelvis gensidig undervisning – og en anden klasse, der ikke modtager denne behandling). Den anden er ændringer over tid (før – efter). De bliver fortolket forskelligt, og det er et empirisk spørgsmål, om der er forskel på dem. De fleste af effekterne i *Visible Learning* er baseret på gruppesammenligninger, og effekten af de to sammenligningsmetoder var ikke meningsfuld – det vil sige, at den gennemsnitlige fremgang var lig med fremgangen i gruppesammenligningen. Det ville være muligt i den næste udgave at identificere, hvilken metode der bliver brugt, men med hensyn til allerede foretagne analyser vil det ikke have indflydelse på de generelle budskaber.

2. Bør vi kun bruge de høje effekter og udelukke de lave?

En fejllæsning af VL er, at kun de høje effekter spiller en rolle, og at vi burde afholde os fra at bruge de påvirkninger, der har en virkning tæt på nul. Dette er alt for forenklet. Nogle af de lave effekter burde interessere underviserne, og en undersøgelse af, hvorfor nogle effekter er lave, kan være afgørende for for-

ståelsen af, hvordan deres virkning kan forøges. For eksempel: At forstå, hvorfor effekterne af klassestørrelser er små (men positive) kan hjælpe os til at forstå, hvordan vi kan forøge virkningen af at reducere klassestørrelsen (Hattie 2007). At forstå, hvorfor effekten af lærernes faglige viden er så lav, er essentiel for at forøge (hvilket vi burde) virkningen af denne påvirkning, og vores team arbejder stadig med dette problem.

På samme måde kan det, der ligner små effekter, være værdifulde i skolesituationer – fordi de kan vise, at en intervention går i den rigtige retning. De kan også vise, at dybere processer måske er ved at ændre sig, og de kan vise, at mere tid, højere grader af implementering eller justeringer er nødvendige. Blot fordi en effekt ikke er $> 0,40$ betyder ikke, at den ikke er noget værd. For eksempel har de tre metaanalyser af internetbaseret læring alle fundet effekter omkring $d = 0,20$ (Olson og Wisner 2002, $d = 0,24$; Sitzman, Kraiger, Stewart og Wisner 2006, $d = 0,15$; Mukawa 2006, $d = 0,14$). Når man derfor implementerer et internetbaseret læringsprogram, kan den generelle effekt af disse tre metaanalyser ($d = 0,18$, baseret på 136 effekter, 22.554 elever) måske være et mere hensigtsmæssigt gennemsnit at sammenligne med end det generelle $d = 0,40$.

3. Studierne i synlig læring

Det blev bemærket i *Visible Learning*, at spørgsmålet om kodning af metaanalyserne efter kvalitet ikke ville blive behandlet i dybden. Dette er blevet misfortolket som en påstand om, at jeg ikke er interesseret i kvaliteten. Det er ikke rigtigt. I teksten blev det bemærket, at andre har udført denne forskning, og der synes ikke at være megen grund til at gentage disse påstande. Det er et empirisk spørgsmål, om kvaliteten af studiet er en moderator. Som forklaret i *Visible Learning*: Lipsey og Wilson (1993) sammenfattede for eksempel 302 metaanalyser inden for psykologi og under-

visning og fandt ingen forskelle mellem studier, der kun inkluderede randomiserede kontra ikke-randomiserede studiedesigns ($d = 0,46$ versus $d = 0,41$) eller mellem studier af høj kvalitet ($d = 0,40$) og lav kvalitet ($d = 0,37$). Der var en bias opad i de publicerede studier ($d = 0,53$) sammenlignet med de ikke-publicerede studier ($d = 0,39$), selvom udsnitsstørrelsen ikke var relateret til effektstørrelsen ($d = -0,03$). Sipe og Curlette (1996) fandt ingen relation mellem den generelle effektstørrelse af 97 metaanalyser ($d = 0,34$) og udsnitsstørrelse, antallet af kodede variable og typer af forskningsdesigns samt en let forøgelse for publicerede ($d = 0,46$) versus upublicerede ($d = 0,36$) metaanalyser. Der er en undtagelse, som kan forudsiges ud fra statistiske principper: Hvis effektstørrelserne er tæt på nul, er sandsynligheden for høj tillid til denne effekt sandsynligvis relateret til udsnitsstørrelsen (se Cohen 2013). Målet bør være at sammenfatte alle mulige studier uanset deres design og derefter undersøge, om kvalitet er en moderator for de endelige konklusioner. Hvor det var hensigtsmæssigt, blev det bemærket, når kvalitet var en moderator. Kritikerne af forskningen må have fejllæst udtalelserne om kvalitet.

4. Læringsbegrebet i VL

Kritikken går på, at *Visible Learning* opsummerer alt, hvad der er blevet gjort dårligt i de sidste 150 år. Alt for mange lærere er alt for docerende, alt for fokuserede på overfladeresultater, testene er af dårlig kvalitet, og de forhold, lærerne arbejder under, er ikke fremmende for bedre resultater. Alt dette kan faktisk godt være tilfældet, men evidensen viser stadig, at der findes fremragende undervisning og skoler overalt omkring os. Evidensen viser, at rigtig mange lærere systematisk opnår mere end et års vækst for et års input, og synlig læring tager udgangspunkt i denne erkendelse og beder om, at disse lærere bliver individuelt anerkendt og værdsat, og at de kommer i centrum af opbygningen af en koalition af succes i vores

system. Det er en kendsgerning, som bemærket i konklusionen i *Visible Learning*, at alt for mange resultater (ud fra lærerproducerede og standardiserede målinger) er på overfladeniveau – men det er en afspejling af, hvad der foregår i skolerne. Måske er dette uhensigtsmæssigt, men den sædvanlige respons på dette er at hylde dyb læring. Som vores syntese af læringsstrategier viser, er det, der er brug for, et passende forhold mellem overflade og dybde for at overføre læring. Dette forhold er i høj grad en funktion af, hvad der bliver undervist i (novicer har brug for mere overflade i begyndelsen for så at gå i dybden i det rigtige øjeblik og derefter lære at overføre og udvide forståelsen). Dette forhold mellem at vide, hvornår man skal holde op med at lære "mere" og begynde at relatere og udvide, er en del af undervisningens kunst og videnskab.

Konklusioner

Uddannelsesforskningens tilstand er sund – vi ved en hel masse. Der findes en undervisningspraksis. Vi har robuste skuldre at stå på. Vi er måske ikke enige i fortolkningerne, men der findes fascinerende og systematiske forskningsresultater. Jeg er for eksempel fascineret af mange af de påstande om, hvad der gør en forskel, som synes så forførende, som gentages så ofte, men som har så lidt evidens til støtte for sig. Spørgsmål om, hvorfor der findes så megen evidens for, at effekterne af nogle af disse påstande (for eksempel klassestørrelse, niveaudeling, forskellige skoleformer) er så små, kan føre i nye retninger. Tag for eksempel den evidens, der viser, at en forbedret faglig viden hos lærerne har tæt på ingen effekt. Hvis man accepterer denne evidens, er det nødvendigt at stille spørgsmålet om, hvorfor denne effekt er så lav. Herefter åbnes muligheden for nye måder at forøge virkningen af denne påvirkning på. At ignorere evidensen og sige, at det må være sandt, at faglig viden spiller en rolle, er ikke at støtte fremskridt inden for feltet.

Der er stadig meget at gøre. For eksempel er mere debat om begrebet "læring" nødvendig, især i skoleårene 0 til 5. Jeg er ved at være færdig med en meta-syntese om læringsstrategier, som direkte behandler dette spørgsmål, og den er baseret på en sondring mellem overfladelæring, dyb læring og overførsel af læring. Meget af *Visible Learning* er baseret på, hvad der sker i de faktiske klasseværelser, og som bemærket i bogen er alt for meget af dette overfladelæring – ikke kun, fordi det er det, der bliver vurderet, men fordi det er det, der synes at blive tillagt værdi i vores skoler.

På samme måde er evidensen i *Visible Learning* baseret på kvantitative mål for, hvad der sker i klasseværelser, og for læring. Der findes også en righoldig kvalitativ litteratur, og en af de mere spændende udviklinger over det seneste tiår har været den systematiske syntese af kvalitativ forskning (se Suri 2014). Sådanne syntese kan føje en værdifuld rigdom til evidensen om klasseværelser og læring. Der er endvidere mange andre resultater af skolen end præstationer, og det er spændende at høre om moderne forskning af teams, der syntetiserer forskning i forbedrede holdninger, motivation og andre affektive resultater. Som nævnt er der nu bøger, der syntetiserer forskning, som specielt beskæftiger sig med elevernes læringsbehov (Swanson 1999; Mitchell 2014).

Meget vigtig er opsummeringen af synlig læring som "kend din virkning" – og dette gør det nødvendigt at stille de tre afgørende spørgsmål, der burde debatteres i enhver skole og ethvert skolesystem: Hvad mener vi med virkning? Hvad er den ønskede størrelse af denne virkning (hvad betyder det for eksempel at sige "mindst et års fremgang for et års input"?). Og hvad er omfanget af denne ønskede virkning (hvor mange elever opnår eksempelvis dette års fremgang for et års input?). Disse spørgsmål afgøres ikke af forskningssynteser, men af vores offentlige demokrati. Hvis vi ikke stiller og på en eller anden kollektiv måde opnår enighed om disse spørgsmål om virkning, vil lærerne fortsætte med at maksimere, hvad de ser som svarene, og det vil betyde, at eleverne primært vil lære som en funktion af den enkelte lærers system af overbevisninger.

Endelig leverer synlig læring, ligesom al anden forskning, retninger og fortolkninger og sigter på at forfine det bedste sæt spørgsmål til fremtidige studier. Om de er relevante for din skole, i dit land og for dine elever er et spørgsmål om forskning – sandsynligheden er (mener jeg), at fortolkningerne af evidensen i synlig læring er relevante for de fleste vestlige demokratier, men også det er en hypotese, der kan testes i.

Litteratur

- Alexander, R. (2011): "Evidence, rhetoric and collateral damage: the problematic pursuit of 'world class' standards". I *Cambridge Journal of Education*, 41(3), s. 265-286.
- Borenstein, M.; Hedges, L.V.; Higgins, J.P. & Rothstein H.R. (2011): *Introduction to meta-analysis*. John Wiley & Sons.
- Clinton, J.M. & Hattie, J.A.C. (2014): "Education and Empowerment Evaluation". I Fetterman, D.; Kaftarian, S. & Wandersman, A. (red.): *Empowerment Evaluation: Knowledge and Tools for Self-assessment, Evaluation Capacity Building, and Accountability*. Sage.
- Cohen, J. (2013): *Statistical power analysis for the behavioral sciences*. Academic press.

- Cooper, H.; Hedges, L.V. & Valentine, J.C. (red.). (2009): *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.
- Cronbach, L.J. & Snow, R.E. (1977): *Aptitudes and instructional methods. A handbook for research on interactions*. Irvington.
- Gan, M.J.S. & Hattie, J.A.C. (2014): "Prompting secondary students' Use of Criteria, Feedback Specificity and Feedback Levels during an Investigative task". I *Instructional Science*, 42(6), s. 861-878.
- Glass, G.V. (1976): "Primary, secondary, and meta-analysis of research". I *Educational researcher*, 5(10), s. 3-8.
- Hansford, B.C. & Hattie, J.A. (1982a): "Self-measures and achievement: Comparing a traditional review of literature with a meta-analysis". I *Australian Journal of Education*, 26, s. 71-75.
- Hansford, B.C. & Hattie, J.A. (1982b): "The relationship between self and achievement/ performance measures". I *Review of Educational Research*, 52, s. 123-142. (Finalist til American Educational Research Associations pris Palmer Johnson Award for excellence in AERA publications.)
- Harks, B.; Rakoczy, K.; Hattie, J.; Besser, M. & Klieme, E. (2013): "The effects of feedback on achievement, interest, and self-evaluation: The role of feedback's perceived usefulness". I *Educational Psychology*, 24(4), s. 269-290.
- Hattie, J.A.C. (1984): "Decision criteria for assessing unidimensionality: An empirical study". I *Multivariate Behavioral Research*, 19, s. 49-78.
- Hattie, J.A.C. (1985): "Methodology review: Assessing unidimensionality of tests and items". I *Applied Psychological Measurement*, 9, s. 139-164.
- Hattie, J.A.C. (1987): "Identifying the salient facets of a model of student learning: A synthesis of meta-analyses". I *International Journal of Educational Research*, 11, s. 187-212.
- Hattie, J.A.C. (1992): "Towards a model of schooling: A synthesis of meta-analyses". I *Australian Journal of Education*, 36, s. 5-13.
- Hattie, J.A.C. (2007): "The paradox of reducing class size and improved learning outcomes". *International Journal of Education*, 42, s. 387-425.
- Hattie, J.A.C. (2009): *Visible learning. A synthesis of 800+ meta-analyses on achievement*. Routledge.
- Hattie, J.A.C. (2012): *Visible learning for teachers. Maximizing impact on achievement*. Routledge. På dansk: *Synlig læring. For lærere*. Dafolo (2013).
- Hattie, J.A.C. (2013): "The power of feedback in school settings". I Sutton, R. (red.): *Feedback. The handbook of criticism, praise, and advice*. Peter Lang.
- Hattie, J.A.C. & Brown, G.T.L. (2008): "Technology for school-based assessment and assessment for learning: Development principles from New Zealand". I *Journal of Educational Technology Systems*, 36(2), s. 189-201.
- Hattie, J.A.C. & Clinton, J.M. (2011): "School leaders as evaluators". I Reeves, D. & Hattie, J.A.C.: *Activate. A leader's guide to people, practices and processes*. The Leadership and Learning Center.
- Hattie, J.A.C.; Masters, D. & Birch, K. (2015): *Visible Learning in Action*. Routledge. Udkommer på dansk under titlen *Synlig læring i praksis* den 1. maj 2016 (Dafolo).
- Hattie, J.A.C. & Timperley, H. (2007): "The power of feedback". I *Review of Educational Research*, 77(1), s. 81-112. På dansk: "Styrken ved feedback". I Andreassen, R. m.fl.: *Feedback og vurdering for læring*. Dafolo (2013).
- Hattie, J.A.C. & Yates, G. (2014): *Visible Learning and the Science of how we Learn*. Routledge. På dansk: *Synlig læring og lærigens anatomi*. Dafolo (2014).
- Hedges, L.V. & Olkin, I. (2014): *Statistical method for meta-analysis*. Academic press.
- Huedo-Medina, T.B.; Sánchez-Meca, J.; Marin-Martinez, F. & Botella, J. (2006): "Assessing heterogeneity in meta-analysis: Q statistic or I² index?" I *Psychological methods*, 11(2), s. 193.
- Kennedy, M.M. (2008): "Contributions of qualitative research to research on teacher qualifications". I *Educational Evaluation and Policy Analysis*, 30(4), s. 344-367.
- Lipsey, M.W. & Wilson, D.B. (1993): "The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis". I *American psychologist*, 48(12), s. 1181.

- Lipsey, M.W. & Wilson, D.B. (2001): *Practical meta-analysis* (49). Sage publications.
 - Mitchell, D. (2014): *What really works in special and inclusive education: Using evidence-based teaching strategies*. Routledge. På dansk: *Hvad der virker i inkluderende undervisning. Evidensbaserede undervisningsstrategier*. Dafolo (2014).
 - Mukawa, T.E. (2006): "Seven principles for good practice and effective online instruction in higher education". I *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, 1, s. 236-240.
 - Olson, T. & Wisner, R. A. (2002): "The effectiveness of web-based instruction: An initial inquiry". I *The International Review of Research in Open and Distributed Learning*, 3(2).
 - Popper, V.K. (1968): *The Logic of Scientific Discovery*. Hutchinson.
 - Rowland, C.A. (2014): "The effect of testing versus restudy on retention. A meta-analytic review of the testing effect". I *Psychological bulletin*, 140(6), s. 1432-1463.
 - Schulze, R. (2004): *Meta-analysis-A comparison of approaches*. Hogrefe Publishing.
 - Senge, P.M. (1990): "The leader's new work. Leading organizations". I *Sloan Management Review*, 32(1) s. 7-23.
 - Sipe, T.A. & Curlette, W.L. (1996): "A meta-synthesis of factors related to educational achievement: a methodological approach to summarizing and synthesizing meta-analyses". I *International Journal of Educational Research*, 25(7), s. 583-698.
 - Sitzmann, T.; Kraiger, K.; Stewart, D. & Wisner, R. (2006): "The comparative effectiveness of web-based and classroom instruction: A meta-analysis". I *Personnel Psychology*, 59(3), s. 623-664.
 - Suri, H. (2014): *Towards methodologically inclusive research syntheses. Expanding possibilities*. Routledge.
 - Suri, H. & Hattie, J.A.C. (2013): "Meta-Analysis and Research Synthesis in Education". I Meyer, L. (red.): *Oxford Bibliographies in Education*. Oxford University Press.
 - Swanson, H.L. (1999): *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. Guilford Press.
-