

Från lärare till skolor -

om att införa formativ bedömning i större skala



Siobhan Leahy & Dylan Wiliam

©Leahy S. & William D. 2009

Originalalets titel: *From Teachers to Schools:
Scaling up Professional Development for
Formative Assessment*

Paper presented at the AERA annual
conference 2009

Översättning: Sara Hjelm (2014)

Inledning

Kirurgen Atul Gawande (2007) har påpekat att inom högt krävande medicinska områden, som exempelvis kirurgi, har framstegen varit större än inom skenbart "enklare" områden som grundläggande hygien. Som exempel nämner han handtvätt efter kontakt med patienter, där efterlevnaden på många sjukhus är under 50 %, trots att det är allmänt känt att det krävs en efterlevnad på över 95 % för att styra spridningen av resistenta former av Stafylokocker som VRSA och MRSA (Gawande, sid 15).

Problemet är inte okunnighet eller avsiktlig olydnad. En strikt rutin är angiven för handtvätt, men som Gawande påpekar:

"... nästan ingen följer den här rutinen. Den verkar helt omöjlig att införa. På morgonronden kontrolleras 20 patienter under en timme. Sjuksköterskor på intensivvårdsavdelningar har vanligen ett liknande antal kontakter med patienter som kräver handtvätt. Även om man pressar hela processen till en minut per patient, så blir det fortfarande en tredjedel av personalens tid som går åt till bara handtvätt. Sådan frekvent handrengöring kan också irritera skinnet, vilket leder till hudinflammation, vilket i sin tur ökar antalet bakterier (sid. 18).

Vid en analys verkar lärares och sjukvårdsklinikers praktik ha många gemensamma drag. Arbetet sker i så snabbt tempo att det finns liten tid för reflektion och beteenden på arbetsplatsen grundas lika mycket på vana som på något annat.

Det skulle inte göra så mycket om vanorna var de mest effektiva för att öka elevers kunskaper, men det finns gott om bevis för att en stor del av lärares praktik är suboptimal och att betydande förbättringar av elevers resultat är möjliga om lärares klassrumspraktik förändras.

I den här rapporten beskriver vi våra ansträngningar att utveckla metoder som stödjer förändring av lärares undervisning - genom fokus på formativ bedömning och med särskilt fokus på att finna modeller där principerna kan införas i större skala, till exempel i 300 000 klassrum i England eller över 2 miljoner klassrum i USA.

I de följande två avsnitten sammanfattar vi kort den forskning som lett fram till vårt ställningstagande att formativ bedömning är den effektivaste metoden för att förändra lärares praktik på ett sådant sätt att det med största sannolikhet gynnar de studerande och också varför vi valt just *"teacher learning communities"* som ett verktyg för förändringsarbetet.

I de senare avsnitten beskriver vi en modell för professionellt utvecklingsarbete som bidrar till att skapa skolbaserade *"teacher learning communities"* och som hjälper lärare att utarbeta egna modeller av formativ bedömning.

Vi avslutar med en fallstudie av genomförandet i ett av Londons storstadsdistrikt.

Argumenten för formativ bedömning

Beläggen för att formativ bedömning är en kraftfull hävstång för att lyfta elevprestationer har stadigt ackumulerats det sista kvartssekle. Under den här tiden har åtminstone 15 betydande forskningsöversikter, som tillsammans sammanfattar flera tusen undersökningar, dokumenterat

effekten av bedömning i klassrumssituationen, för elever och för studerande (Fuchs & Fuchs 1986; Natriello, 1987; skojare, 1988; Bangert-Drowns, Kulik, Kulik & Morgan, 1991; Dempster, 1991, 1992; Elshout-Mohr, 1994; Kluger och DeNisi, 1996; Black & Wiliam, 1998; Nyquist, 2003; Brookhart, 2004; allal & Lopez, 2005; Köller, 2005; Brookhart, 2007; Wiliam, 2007; Hattie & Timperley, H 2007; Shute, 2008).

Medan många av översikterna dokumenterat negativa effekter av vissa bedömningsmetoder, så visar de också att utnyttjad på lämpligt sätt har bedömning en betydande potential att förbättra elevers resultat.

Utifrån ett tidigt arbete av Scriven (1967) och Bloom (1969), har varit vanligt att beskriva användningen av utvärdering för att förbättra studerandes lärande som "*formativ bedömning*", även om också termen "*bedömning för lärande*" nyligen har blivit vanligare. I USA har ofta begreppet "*bedömning för lärande*" felaktigt tillskrivs Rick Stiggins (2002), även om Stiggins själv alltid tillskrivit den författare i Storbritannien. I själva verket verkar den första förekomsten av *formativ bedömning* i den nuvarande betydelsen vara i ett bidrag till den årliga konferensen för *Association for Supervision and Curriculum Development* (James, 1992) medan begreppet tre år senare användes som titel på en bok (Sutton, 1995). Den första användningen av begreppen "*bedömning för lärande*", i kontrast till "*bedömning av lärande*", verkar kunna hänföras till boken *Assessment: a teacher's guide to the issues* (Gipps & Stobart, 1997), där de två termer är titlarna på första respektive andra kapitlen. Skillnaden spreds till en vidare publik i en handbok för beslutsfattare från Assessment Reform Group 1999 (Broadfoot, Daugherty, Gardner, Gipps, Harlen, James & Stobart, 1999).

Wiliam (2009) sammanfattar några av de definitioner för formativ bedömning (och bedömning för lärande) som har lagts fram under åren, och föreslår att den mest heltäckande definitionen är den som används av Black och Wiliam (2009):

Klassrumspraktiken är *formativ* om bevis för elevresultat tas fram, tolkas och används av lärare, elever eller deras jämlingar för att bestämma om beslut om nästa steg i undervisningen sannolikt är bättre eller bättre grundat än de beslut som de skulle ha tagit i avsaknad av den bevisning som tagits fram (sid. 9).

I en kommentar till den här definitionen, betonar Black och Wiliam att fokus för betydelsen av ordet **beslut** utgör en kompromiss mellan att basera definitionen på **avsikt** (vilket skulle vara en mycket vid definition, där nästan all datainsamling kan ses som formativ) och att basera den på **resultat** (vilket skulle vara en mycket snäv definition på grund av själva lärandets oförutsägbara natur).

De påpekar också, att medan sådana beslut tas regelbundet av läraren, så kan samtidigt eleverna och deras klasskamrater vara inblandade i besluten. Faktiskt är det, när allt kommer omkring, precis så som Monique Boekaerts arbete visar - att om de studerande inte själva väljer ökat lärande istället för personligt välbefinnande så kan man bara förvänta sig att en obetydlig mängd lärande kommer att ske (Boekaerts, 2006).

Ett tredje påpekande är att definitionen är probabilistisk, återigen på grund av det oförutsägbara i lärande, och det fjärde påpekandet är att formativ bedömning inte behöver ändra undervisningen för att vara formativ - utan bara bekräfta att de föreslagna åtgärderna verkligen är de mest lämpliga.

Den allmänna slutsatsen är att inom en rad olika skolämnen, i olika länder och för elever i olika åldrar, förefaller användningen av formativ bedömning vara associerad med betydande förbättring av lärandet. Att beräkna hur stora vinsterna är kan vara svårt, eftersom de flesta undersökningar verkar ignorera det faktum att resultatet av mätningarna skiljer sig åt beroende på typen av undervisning (Wiliam, 2009), men slutsatsen att användning av formativ bedömning kan öka studerandes lärande med någonstans mellan 50 och 100 procent verkar rimlig.

Det här antyder i sin tur att formativ bedömning sannolikt är ett av de mest effektiva sätten och kanske det mest effektiva för att förbättra elevers resultat (till exempel beräknar Wiliam & Thompson 2007, att det skulle vara 20 gånger mer kostnadseffektivt än gängse handlingsprogram för minskning av klasstorlek).

De omfattande bevisen för formativ bedömnings potentiella kostnadseffektivitet och funktion som hävstång för skolutveckling har, som väntat, väckt stor uppmärksamhet. Ett antal testförlag har gjort vad de kallar "*Formative assessment system*" eller "*Benchmark assessments*". Dessa inkluderar **MAP** producerat av Northwest Evaluation Association (NWEA), **Focus on Standards™/Instructional Data Management System™** från ETS, **Homeroom™** från Princeton Review, **Benchmark Tracker™/SkillWriter™** och **Stanford Learning First™** producerade av Harcourt Assessment, **Prosper™** producerat av Pearson Assessments och en mängd andra liknande produkter.

Vanligtvis gäller det system för bedömning av elevers utveckling med regelbundna intervall (oftast var fjärde till var nionde vecka) och de tillhandahåller rapporter som identifierar elever eller särskilda aspekter av kursinnehållet som kräver särskild uppmärksamhet. Medan vissa bara hänvisar till existerande litteratur om formativ bedömning som bevis för sin effektivitet, har andra utgivare ägnat sig åt egen forskning om effekterna. Som ägare av "*Fokus on Standards™*", har ETS undersökt effekterna av programmet och konstaterar att även om det kan ha en betydande effekt under särskilda omständigheter (till exempel när överensstämmelsen mellan kursplanen och undervisningens innehåll är dålig) så tycks de allmänna effekterna vara begränsade (Goe & The Bridgeman, 2006).

När det gäller den ovan diskuterade definition som föreslås av Black och Wiliam, kan sådana system kan vara formativa, i den meningen att de kan lägga fram bevis för elevresultat som i sin tur kan användas för att fatta bättre beslut om undervisningen än vad som hade varit möjligt utan. Men eftersom få om än någon av de 15 studier som nämndes inledningsvis behandlar liknande "lång-cyklig" formativ bedömning, så kan man på grundval av befintlig forskning inte dra slutsatsen att sådana periodiska utvärderingar sannolikt ger betydande effekter på elevernas resultat.

Formella system för att regelbundet testa elever kan visst spela roll att när det gäller att övervaka elevernas framsteg effektivt och vissa sätt att spåra studerandes framsteg på medellång sikt och att vidta åtgärder för att åtgärda eventuella identifierade problem kan synas vara en avgörande del av vilken omfattande bedömning som helst. Men det är minst sagt oärligt och eventuellt lögnaktigt att hävda att forskningen ger belägg för sådana systems effektivitet. Det gör den helt enkelt inte (Popham, 2006; Shepard, 2007). Därmed inte sagt att sådana belägg inte kan komma i framtiden, det kan de mycket väl - men mycket få har samlats in i dagsläget.

Samma sak kan sägas om det som kallas "*gemensam formativ bedömning*" eller "*fortlöpande bedömning*", definierat av DuFour, DuFour, Eaker, m.fl. (2005) som:

... en bedömning, vanligtvis skapad av ett team av lärare som samarbetar med ansvar för samma klass eller kurs. Gemensam formativ bedömning administreras ofta under året för att identifiera (1) enskilda studenter som behöver extra tid och stöd för lärande, (2) de mest effektiva pedagogiska strategierna för att hjälpa elever att förvärva de avsedda kunskaperna och färdigheterna, (3) svårigheter i innehållet – områden där elever i allmänhet har bekymmer med att uppnå den avsedda målnivån och (4) målsättning för enskilda lärare och arbetslag (sid. 214).

Återigen, medan sådana bedömningar helt klart har en viktig roll att spela när det gäller att anpassa undervisningen till önskad standard, att tjäna som fokus för en professionell dialog och att stödja god förvaltning och tillsyn, så är samtidigt bevisen svaga för att sådan formativ bedömning i medel långa cykler ger effekter på elevers resultat.

Däremot finns det starka bevis för att det som Wiliam och Thompson (2007) kallar "kort-cyklisk" formativ bedömning kan ha långtgående effekt på elevresultat. Yeh (2006) sammanfattar ett antal studier som visar att vad han kallar "snabb formativ bedömning" (bedömningar från två till fem gånger per vecka) avsevärt kan förbättra elevers inläring. På en ännu kortare tidsskala (Black, Harrison, Lee, Marshall och Wiliam, 2003) beskrivs hur de gav stöd till en grupp på 24 matematik- och naturvetenskapslärare för att utveckla användningen av formativ bedömning "i-stunden" och fann att även när den mäts genom externt framtagna och poängsatta uppgifter i standardiserade prov, så var vinsterna i elevresultat betydande, och motsvarade en ökning av elevers lärande med 70 procent (Wiliam, Lee, Harrison & Svart, 2004). Andra liknande insatser har lett till motsvarande effekter (Hayes, 2003; Clymer & Wiliam, 2006/2007).

Därför verkar det rimligt att de effekter som litteraturen visar är möjliga, verkligen kan uppnås i reella klassrum, även om resultaten mäts med externt mandat och standardiserade tester. Det är mycket mindre klart hur man skall uppnå dessa effekter i större skala – i över 300 000 klassrum i England, eller över 2 miljoner klassrum i USA.

Att utforma för större skala

När det gällde att utforma metoder som stödjer införandet av formativ bedömning i ett stort antal klassrum, antog vi och våra kollegor på Educational Testing Service en design begränsad av begreppet "*skalbarhet i princip*". Med detta menade vi att införandet inte behövde vara i större skala i början, men om någon aspekt inom rimliga gränser inte kunde införas i större skala så uteslöts den.

En andra begränsning var att det skulle bli en gemensam modell för *hela* skolan. En av de mest överraskande lärdomarna från vårt arbete med skolor under 20 år är hur uppdelade fortbildningsarrangemang för lärare är, framför allt i gymnasieskolan. Det är ganska vanligt att hitta matematiklärare sysselsatta med en uppsättning kompetensutveckling, medan naturkunskapslärare och lärare i samhällsvetenskapliga ämnen gör något helt annat. Bortsett från att det blir svårt och förvirrande för eleverna, så gör skillnaderna i synsätt det mycket svårare att skapa en sammanhängande dialog i skolan kring undervisning och lärande. Och med en enda modell för hela skolan som mål, insåg vi också att vi måste ta hänsyn till särdragen för ålder och ämne. Undervisning av femåringar är inte samma sak som undervisning av tioåringar och att studera matematik inte är samma sak som att studera historia.

Vi blev också medvetna om att en modell för effektiv, utbyggbar fortbildning för lärare skulle behöva klara en svår balansgång mellan två motstridiga krav. Den första var behovet av att se till

att modellen var tillräckligt flexibel för att anpassas till lokala förhållanden, inte bara för att lyckas, utan också för att kunna utnyttja alla tillgångar som finns i det lokala sammanhanget och som skulle kunna förbättra en introduktion. Det andra var att säkerställa att modellen var tillräckligt stram för att bevara tillräcklig trohet mot ursprungsmodellen och ge en rimlig försäkran om att en eventuell modifiering inte ledde till att åtgärderna genomgick en "förgörande mutation" (Haertel, citerad i Brown & Campione, 1996).

För att lösa problemet, antog vi uttryckligen ett ramverk med titeln "Stram men fri":

Formulering *Stram men fri* kombinerar en överdriven trohet mot centrala principer för formen (den "strama" delen) och med utrymme för anpassning till behov, resurser, begränsningar och egenheter som finns i en skola eller ett distrikt (den "fria" delen), men bara om det inte strider mot teorin för interventionen. (Thompson & Wiliam, 2008 sid. 35; originalets betoning).

En utförligare redogörelse för tillämpningen av ramverket "Stram men fri" vid utformning av ett program för kompetensutveckling som stöd för lärare i deras användning av formativ bedömning, finns i Thompson och Wiliam (2008). Av särskild betydelse är att vår modell styrdes av en rad principer som tidigare arbete hade visat vara viktiga när det gäller hjälp till lärare i utvecklingen av formativ bedömning i undervisningen - i synnerhet valfrihet, flexibilitet, små steg, ansvar och stöd (Wiliam, 2006).

Valfrihet

Det är ett vanligt antagande att lärare ska arbeta med att utveckla de svagaste aspekterna av sin praktik för att förbättra och för en del lärare kan de här aspekterna i själva verket vara så svaga att de bör prioriteras vid kompetensutveckling.

Men vår erfarenhet är att sådant som varit till störst nytta för eleverna är det som för de flesta lärare leder till att de blir ännu mer experter på sina styrkor.

Tidigt i arbetet med formativ bedömning för lärare i England (Black, Harrison, Lee, Marshall & Wiliam, 2003), var en av dem, Derek (alla namn på lärare, skolor och kommuner som nämns i detta dokument, är pseudonymer) redan ganska duktig på att leda diskussioner i helklass, men han var samtidigt intresserad av att förbättra det ytterligare. En kollega till honom på samma skola, Philip, hade en mycket lägre profil i klassrummet och var mycket mer intresserad av att hjälpa eleverna utveckla färdigheter i självvärdering och kamratbedömning. Både Derek och Philip är utomordentligt skickliga läkare, bland de bästa vi sett, men att få Philip att arbeta med modeller för kontrollfrågor, eller Derek att arbeta med kamratbedömning och självbedömning trodde vi inte skulle gynna eleverna lika mycket som att stötta dem att själva utveckla sin expertis. Dessutom har vi funnit att när lärare själva fattar beslut om vad de vill prioritera för sin egen professionella utveckling, så är det mer troligt att de "får det att fungera". I traditionella "top - down" -modeller för kompetensutveckling, får läraren pröva idéer i sin praktik, men ofta slutar det med att de skyller på ledaren av kompetensutvecklingen om de nya metoderna misslyckas i klassrummet (t.ex. "*jag har prövat vad du sa till mig och det fungerade inte*").

Om valet av metoder och aspekter av utveckling däremot görs av läraren, så delas ansvaret för att garantera ett effektivt genomförande.

Flexibilitet

Lärare behöver flexibiliteten för att kunna ändra eller modifiera tekniker i formativ bedömning som introduceras - så att de passar in i sammanhanget i det egna klassrummet (Ginsburg, 2001). Risken med detta är att en lärare tillåts modifiera en idé tills den inte längre är verksam (se exemplet på "förgörande mutation" ovan). Vad som behövs är därför ett sätt att tillåta flexibilitet, men samtidigt begränsa den tillräckligt, så att ändringarna inte försämrar ursprungsidéns effektivitet i onödan. Vår lösning var att klargöra skillnaden mellan strategier och tekniker för strukturerad bedömning. För att definiera strategier, började vi med att särskilja tre viktiga processer i formativ bedömning:

- Identifiera var eleverna befinner sig i sitt lärande
- Identifiera vart de ska
- Identifiera vilka åtgärder som krävs för att komma dit

Med utgångspunkt från lärares och elever respektive roller gav dessa processer den tabell med nio celler som visas i figur 1 och i vilken de fem "strategierna" för formativ bedömning finns komprimerade (William & Thompson, 2007).

Var och en av de fem "strategierna" ger fokus för en rad aspekter i lärares praktik. Eller med andra ord ger de en utgångspunkt för reflektion över av ett antal omfattande frågor som rör undervisning, som till exempel innehåll, psykologi och pedagogik.

För var och en av de fem nyckelstrategierna identifierades ett antal specifika klassrumsrutiner, så kallade "tekniker". Teknikerna har form av "validerade metoder" som överensstämmer med forskningen om formativ bedömning och lärare är fria att välja vilken av dessa tekniker de vill införa (vilket därmed ger valfrihet - se ovan). Genom att förankra tekniker till (minst) en av de fem nyckelstrategierna, fanns en möjlighet för läraren att modifiera, men att ändå behålla en rimlig garanti för överensstämmelse med forskningsunderlaget.

Figur 1. Aspekter på formativ bedömning

	VART ÄR ELEVEN PÅ VÄG?	VAR BEFINNER SIG ELEVEN?	HUR SKA ELEVEN TA SIG VIDARE?
LÄRARE	Tydliggöra mål och kunskapskrav (1)	Organisera effektiva diskussioner och aktiviteter som utvecklar och som synliggör elevers lärande (2)	Ge feedback som hjälper eleverna framåt (3)
KAMRATER	Förstå och instämma i vilka målen för inläring och vilka kriterierna för framgång är (1)	Aktivera eleverna som resurser För varandras lärande (4)	
ELEVER	Förstå målen för inläring och kriterier för framgång (1)	Aktivera eleven som ägare av sitt eget lärande (5)	

Små steg

Vilken modell av fortbildning man än använder, så måste man acceptera att lärare lär sig långsamt. I synnerhet när det gäller att förändra sin praktik. I motsats till ny kunskap, så måste ny praktik integreras i lärarens befintliga rutiner för att bli beständig och det tar tid. Många fortbildare har erfarenhet av att först uppmuntra till att pröva nya idéer, sen se dem användas när de besöker undervisningen och slutligen, strax efteråt, få veta att lärarna återgått till tidigare arbetssätt.

En del tolkar sådan regression som motstånd från lärarnas sida - orsakat av lojalitet mot de yrkesmässiga vanor som i realiteten är kärnan i den professionella identiteten. Det kan vara *en del* av orsaken till den långsamma förändringstakten och att förändringar inte blir bestående. Men mer betydande enligt vår bedömning, är att hög prestanda i en så komplex domän som undervisning kräver automatisering av en stor del av de saker som lärare gör.

För körskoleeleven verkar växling, att använda av körriktningsvisare och att styra samtidigt ofattbart komplicerat - och det är det också om man försöker genomföra det medvetet. Erfarna bilförare har gjort det här så många gånger att handlingarna blir automatiserade och därmed inte tar upp så mycket av de tillgängliga resurserna för kognitiv bearbetning. Men om någon försöker ändra sitt körsätt, till exempel minska hur mycket de "slirar på kopplingen" - så upptäcker de att automatiserade handlingar är extremt svåra att förändra.

Undervisning är till och med ännu mer extrem än bilkörning i det här avseendet, eftersom alla lärare kommer till yrket med en serie av "manualer" för hur klassrum bör fungera, "inprogrammerade" i medvetandet från den tid när de själva var elever. De här manualerna, som till exempel kräver att eleverna ska räcka upp handen om de kan svaret på lärarens frågor, verkar naturliga, men naturligtvis är de inlärd och i många fall olämpliga (Wiliam, 2005).

Många förändringar i praktik som associeras med formativ bedömning är svåra att genomföra och det är inte bara eftersom det rör sig om vanor. Dessutom motsäger de vitt spridda och starkt rotade föreställningar, till exempel värdet av poängsättning för att motivera elever. Även om lärare är övertygade om värdet av "*enbart kommentarer*" vid rättning, så avskräcks de ofta från att pröva av äldre kolleger - som avfärdar innovationer som modenycker, förespråkade av akademiker i elfenbenstorn som inte vet vad verklig undervisning är. Det är därför vi, även om vi har bråttom att hjälpa lärare att förbättra sin praxis, bör "skynda långsamt".

Stöd och ansvar

De sista två principerna – stöd och ansvar – kan ses som två sidor av samma mynt. Vi har faktiskt, beskrivit dem någon annanstans som ett enda kännetecken för effektiva lärandemiljöer för lärare: stödjande ansvarighet (Ciofalo & Leahy, 2006). Grundtanken är att skapa strukturer som gör lärare ansvariga för att utveckla sin praktik och som också ger dem stöd att göra det.

Man skulle kunna skapa en sådan "stödjande ansvarighet" på ett antal sätt. Ett vore att tilldela varje lärare en personlig coach, men dels vore det dyrt och dels är det ganska otroligt att det finns ett tillräckligt utbud av coacher att tillgå. Kravet på "skalbarhet i princip" ledde till att coachmodellen förkastades och i stället fokuserade vi på idén om *teacher learning communities* (lärargrupper för gemensamt lärande) med arbetsplatsen som bas.

Stöd för formativ bedömning via *Teacher Learning Communities*

Mellan 2003 och 2006, arbetade vi tillsammans med kolleger vid *Educational Testing Service* och utvecklade och sjösatte ett antal modeller för att stötta lärare (för en utförligare redovisning, se Thompson & Goe, 2008; Wylie, Lyon & Goe, 2009; Wylie, Lyon, & Mavronikolas, 2008).

En av våra tidigaste modeller innebar att en av oss (SL) varannan eller var tredje vecka mötte grupper med fyra till sex matematiklärare från *High School*¹ för att diskutera de förändringar de försökte göra i sin praktik. Som ett resultat stod det klart att en tvåveckors cykel inte ger tillräcklig tid för de deltagande lärarna att både planera och hinna genomföra förändringar så att de kan rapportera tillbaka vid nästa möte. Och med sex veckors intervall eller mer mellan mötena tycktes implementeringen förlora i fart.

Det här ledde till beslut om en modell med månatlig cykel för möten och den har ingått i alla våra implementeringar fram till nu. I arbetet med bokstavligen hundratals skolor i ett tiotal distrikt under de senaste fem åren, så har vi inte hittat ett enda bevis som motsäger att cirka fyra veckors intervall mellan möten är det optimala, åtminstone när det gäller sådana förändringar av praktik som är relaterade till formativ bedömning.

Ursprungligen antog vi att skolorna skulle kunna hitta två timmar för månatliga möten, men även om detta inte var något problem i vissa skoldistrikt, så var det helt omöjligt i andra. Därför undersökte vi noga hur mycket mötena skulle kunna minskas i längd. Efter experimenterande med olika möteslängder (inklusive möten så korta som 60 minuter), drog vi slutsatsen att 75 minuter bör vara ett absolut minimum.

Erfarenheter från möten med ett litet antal deltagare ledde oss till slutsatsen att den minsta gruppstorlek som krävs för att ett möte ska vara effektivt är runt åtta personer. Möten med färre än åtta deltagare kräver ofta betydande input från gruppens ledare eller samordnare, särskilt om tillfällig sjukfrånvaro eller andra faktorer minskat storleken på gruppen ytterligare. Även om så intensivt stöd från samordnaren ger en effektiv lärandemiljö för dem som deltar, så skulle en sådan modell sannolikt inte vara möjlig i större skala.

Å andra sidan, om gruppen var mycket större än tolv (vilket ofta var fallet vid våra tidiga projekt i *Cleveland Municipal School District*), så blev det inte tillräckligt med tid för redovisning för varje gruppmedlem. Många har berättat i intervjuer, att det var just det faktum att de visste att de skulle bli skyldiga att redovisa för sina kollegor vad de hade gjort, som fick dem att sätta arbetet med att förändra sin undervisningspraktik före alla andra akuta problem i en lärares liv (Ciofalo & Leahy, 2006).

Förutom riktlinjer för storlek på gruppen och frekvens av möten undersökte vi också i vilken utsträckning det var nödvändigt att lärare hade samma uppdrag (t.ex. undervisning i de lägsta klasserna eller hade samma ämnesspecialisering inom gymnasieskolan). Vår erfarenhet är att lärare uppskattar att arbeta i blandade grupper, med tillgång till idéer från kollegor i olika ämnen eller för olika åldrar. Men vi såg också att lärarna många gånger förkastade förslag från andra gruppmedlemmar med motiveringen att den föreslagna idén inte skulle fungera för det egna ämnet eller nivån. För att balansera sådana spänningar, prövade vi att ge stöd för den egna disciplinen genom att se till att det i blandade grupper fanns ytterligare minst en kollega med samma ålders - eller ämnesspecialisering för varje medlem.

¹High School: elever 14-18 år, nivån motsvarar ungefär svenskt gymnasium

Hittills har vi inga uppgifter som tyder på att en särskild modell för grupp sammansättning skulle vara bättre än någon annan, men vi är samtidigt medvetna om att utrymmet för djupare ämnesdiskussioner sannolikt begränsas om gruppen utgörs av individer med olika specialisering (Grossman, Wineburg & Woolworth, 2000).

En slutlig utformningsdetalj för de månatliga mötena var deras struktur. Som resultat av Shulmans arbete med "signatur pedagogik" för yrken (Shulman, 2004), insåg vi att det kan finnas betydande fördelar med att bestämma en standardstruktur för de här månadsmötena, men också att nyheter betraktas som det viktigaste i de flesta former av lärarfortbildning - för att hålla kompetensen "färsk". Resultatet blir ofta att formen för inläringen får företräde framför innehållet och mycket tid spenderas på att lära sig inläringens struktur, i stället för att ta tid till själva lärandet.

För att lyfta fram själva lärandet och placera inlärningsstrukturen i "bakgrunden", beslutade vi att varje möte skulle arrangeras runt samma sex aktiviteter, som skulle komma i samma ordning varje månadsmöte. Det faktum att varje möte följer samma struktur innebär att deltagarna kommer till mötet och vet vilka roller de ska spela, både vad gäller rapportering tillbaka om de egna erfarenheterna och när det gäller att ge stöd till andra.

1. INTRODUKTION (5 MINUTER)

Agendan för mötet och planeringen för nytt lärande presenteras.

2. STARTAKTIVITET (5 MINUTER)

Deltagarna engageras i en aktivitet som hjälp för att fokusera på det egna lärandet.

3. FEEDBACK (25 MINUTER)

Varje lärare ger en kort rapport om det de åtog sig att pröva under avsnittet "personlig planering" vid föregående möte. Resten av gruppen lyssnar uppskattande och erbjuder sedan sitt stöd för att driva den individuella planeringen framåt.

4. NYTT LÄRANDE OM FORMATIV BEDÖMNING (20 MINUTER)

För att ge varje TLC- möte en del nytt stoff och en stadig ström av nya idéer, har varje möte en aktivitet som introducerar några nya idéer som rör formativ bedömning. Det kan vara en uppgift, en video att titta på och diskutera eller en läsecirkel där lärare diskuterar betydelsen för formativ bedömning utifrån det bokavsnitt de har läst under den senaste månaden.

5. PERSONLIG AKTIONSPLANERING (15 MINUTER)

Den näst sista aktiviteten för varje session innebär att varje deltagare planerar i detalj vad de hoppas att kunna utföra före nästa möte. Det kan inkludera att pröva nya idéer eller det kan helt enkelt vara att konsolidera de tekniker som de redan har experimenterat med. Detta är också en bra tidpunkt för deltagarna att planera för samplanering och lektionsobservationer som de ska genomföra hos varandra. Det är vår erfarenhet att om deltagarna lämnar mötet utan en bestämd dag och tid att observera varandra, så är det mycket mindre sannolikt att observationerna blir av (Maher & Ioffe, 2007).

6. SAMMANFATTNING AV LÄRANDE (5 MINUTER)

De sista fem minuterna av mötet diskuterar gruppen om de har uppnått de avsikter med lärande de ställt upp för mötet. Om de inte har det, finns det tid för gruppen att besluta vad man ska göra åt det.

Från princip till produkt

Så som det sammanfattas ovan ger forskningen om formativ bedömning och principerna för utformning av *teacher learning communities* (TLC) en klar bild av vad som borde hända i klassrummet och av vilka typer av fortbildning som kan hjälpa till att förändra lärares undervisningspraktik i rätt riktning. Men längre än så räcker inte bilden när det gäller att utforma ett koncept som kan implementeras i större skala.

Sedan 2005 har ETS (*Educational Testing Service, New Jersey*) varit inblandade i omfattande pilotprojekt och test av en modell som kallas "*Keeping Learning on Track* TM" och som lanseras senare under 2009². Tidiga versioner av KLT innebar ett fyra dagars träningsprogram (två dagar för samtliga deltagare och ytterligare två dagar för TLC ledare). Materialet till stöd för utbildningen och månadsmötena i TLC uppgick till flera hundra sidor och gav detaljerad vägledning till TLC- ledare om vilka punkter på dagordningen som kunde uteslutas om det var ont om tid. Även om ett så omfattande stödmaterial kan vara nödvändigt i vissa skolor, så finns risken att sådan detaljnivå förmedlar en uppfattning om att den person som leder en TLC måste vara "expert" på formativ bedömning. En av de viktigaste drivkrafterna för de fem principerna ovan var att varje lärare är professionell nog att planera sin egen, personliga, kompetensutveckling och också att, med stöd av likasinnade professionella, kunna genomföra sina planer.

Därför har vi (de två författarna till den här rapporten) sedan 2007 varit inblandade i utvecklingen av en "mini"- uppsättning material, som hjälp till skolor med stöd för lärare i utvecklingen av formativ bedömning i den egna praktiken.

"*Embedding formative assessment*" (Leahy & Wiliam, 2009) är en cd-rom med material för att en skola ska kunna ordna sitt eget endagsseminarium om formativ bedömning (PowerPoint-bilder och en fullständig utskrift för workshop), sju 10-minuters video klipp där en av oss (DW) presenterar forskningsgrunden och några tekniker för formativ bedömning och slutligen material för nio månatliga uppföljningsmöten i TLC- grupp. Materialet till stöd för de nio månadsmötena uppgår till totalt 75 sidor, inklusive dagordningar, åhörarkopior och anteckningar för gruppledaren.

Under tiden som arbetsmaterialet för månadsmötena utarbetades skickades det till att börja med till ett 60-tal skolor som uttryckt intresse för att testa det. Ytterligare utskick gick endast de skolor som lämnat synpunkter på det material de fått.

Den kanske intressantaste upptäckten från den här tidiga fasen i utvecklingsarbetet, var att många skolor la beslag på delar av programmet för att stötta befintlig planering. Trots att de deltagit i seminarier där forskningsgrunden för formativ bedömning diskuterades ganska ingående och där det påvisades att det fanns lite eller ingen forskning till stöd för de andra innovationer inom lärande som drog till sig intresse vid den här tidpunkten (till exempel, "brain gym" eller lärstilar), så rapporterade många skolor att de gillade idén om *teacher learning communities*, men hade beslutat att använda den som stöd för lärare i det som var skolans nuvarande prioritering inom fortbildning (t.ex. differentierad undervisning, individualisering, lärstilar). Naturligtvis är en sådan annektering knappast förvånande, omöjlig att förhindra och kan leda till resultat som är mycket positiva, men det som var förvånande var att de flesta av dem som hade förvandlat modellen till egenkännlighet verkade tro att de genomförde den som det var avsett.

² <http://keepinglearningontrack.nwea.org/index.html>

Utvärderingarna av "*Keeping Learning on Track*™" och "*Embedding formative assessment*" befinner sig båda i sin linda, men utmaningen med att genomföra effektiv fortbildning i större skala är uppenbar för båda programmen.

I resten av den här texten diskuterar vi kortfattat en fallstudie av genomförandet av "*Embedding formative assessment*" i ett skoldistrikt i England. Medan vissa detaljer är unika för det engelska utbildningssystemet, verkar de stora problemen vara exakt desamma som i USA.

Fallstudie - ett skoldistrikt

Cannington³ är ett skoldistrikt i London som har en yta på ca 25 km² och som betjänar en blandad befolkning på cirka 200 000 invånare, med tre gånger så många invånare från etniska minoritetsgrupper som landet som helhet.

I juli 2007 beviljade en välgörenhetsstiftelse medel för ett forskningsprojekt i syfte att höja elevresultaten i matematik, naturvetenskap och moderna språk - ämnen som ledningen i Cannington hade identifierat som prioriterade för utveckling. Möjligheten att undersöka användbarheten i ämnesspecifika *teacher learning communities* i ett helt distrikt gick inte att missa, även om materialet för *Embedding formative assessment* egentligen var avsett för användning i en skola som helhet.

I november 2007, vid ett möte för rektorer i gymnasieskolorna i Cannington, presenterades förslaget att ge stöd till lärare i deras utveckling av formativa metoder för bedömning i klassrummet genom att inrätta tre *teacher learning communities* på varje gymnasieskola: en med fokus på matematik, en med fokus på naturvetenskap och en tredje med fokus på moderna språk.

Medlemmar i projektgruppen deltog i möten med rektorer från Cannington under de följande månaderna för att ge uppdateringar om projektets utveckling och i juli 2008 hölls en serie om tre utbildningstillfällen – ett för varje grupp av ämneslärare från de deltagande skolorna. Antalet lärare från varje skola som deltog vid varje tillfälle visas i tabell 1.

Tabell 1: antalet deltagande lärare per tillfälle

Skola	Matematik	Naturvetenskap	Moderna språk
Ashtree School	1	1	0
Cedar Lodge School	5	1	3
Hawthorne School	4	10	5
Hazeltree School	7	12	2
Larchtree School	1	0	0
Mallow School	6	7	3
Poplar School	11	3	1
Spruce School	7	8	5
Willowtree School	2	5	2
Totalt	44	47	21

³ Namnen på skolorna och på distriktet är fiktiva.

Utbildningsdagen bestod av en kort sammanfattning av forskningen om formativ bedömning, en översikt av de fem "nyckelstrategierna" för formativ bedömning, en introduktion till cirka 30 olika metoder som lärare kan använda för att införa formativ bedömning i sina klassrum och detaljer om skapandet av de tre ämnesspecifika, skolbaserade *Teacher Learning Communities* som skulle sättas samman för att ge kontinuerligt stöd på varje skola. Utbildningen gav också vägledning om rollen som ledare i var och en av de tre grupper som skulle starta på varje skola.

Lärarnas reaktion på utbildningen var extremt positiv och i slutet av dagen verkade deltagarna från sex av de nio skolorna ha en fast plan för genomförande. En skola (Ashtree) beslutade att skjuta upp deltagandet i projektet ett år, och Hazeltree hade tidigare under månaden bestämt sig för att skapa lärargrupper med blandade ämnen snarare än ämnesspecifika, eftersom de ansåg att det var mer i linje med det professionella utvecklingsarbete som redan hade ägt rum på skolan. Eftersom finansieringen hade lämnats speciellt för stöd till lärare i matematik, naturvetenskap och moderna språk, enades man om att detta i praktiken skulle innebära att Hazeltree drog sig ur projektet, även om de fortsatte att ta emot allt material som behövdes till stöd för *Teacher Learning Communities*. Larchtree hade bara skickat en enda lärare (till matematik sessionen), men den berörda läraren verkade övertygad om att hon skulle kunna "bombardera" andra lärare inom matematikavdelningen och möjligen även inom andra ämnen med utbildningsinnehållet.

Även om det inte var möjligt för alla lärare i matematik, naturvetenskap och moderna språk på skolorna att delta i endagsseminarium, så fick alla High School- lärare i dessa ämnen i Cannington ett kort häfte (30 sidor) med beskrivning av de viktigaste principerna för formativ bedömning tillsammans med material för ämnesspecifik anpassning (Hodgen & Wiliam, 2006, Black & Harrison, 2002, Jones & Wiliam, 2007).

För att få till en enkel kommunikationskanal mellan lärarna i skolan och projektet, utsågs sex experthandledare (två för varje ämnesinriktning). Deras huvudsakliga roll var inte att "driva på" genomförandet av programmet i skolorna, utan snarare att svara på frågor från TLC- ledare och skoladministratörer om användningen av materialet. Varje handledare förde en logg över sina kontakter med lärare på skolan, vilket var den huvudsakliga källan till bevis för i vilken utsträckning TLC- grupperna fungerade som avsett.

Med tanke på rektorernas medverkan vid varje steg i processen fram till denna punkt och deras satsning att släppa ett stort antal lärare att delta i inledande workshops, förväntade vi oss att TLC- grupper snabbt skulle etableras, men av olika skäl, var acceptansen extremt ojämn.

När det hade gått sju månader, tolkades handledarnas loggar av två olika bedömare, som var och en uppmanades att betygsätta framstegen för varje TLC- grupp på en skala från 1 (få eller inga framsteg) till 4 (goda framsteg). När bedömningarna - som gjordes oberoende av varandra - jämfördes, så skiljde de sig inte i något fall med mer än en poäng. De poäng som de två bedömare sedan kom överens om framgår av tabell 2.

Tabell 2: Grad av progression för teacher learning communities i respektive skola

Skola	Matematik	Naturvetenskap	Moderna språk
Ashtree School*	-	-	-
Cedar Lodge School	1	1	2
Hawthorne School	2	2	4
Hazeltree School*	-	-	-
Larchtree School	4	1	1
Mallow School	3	1	2
Poplar School	1	3	3
Spruce School	4	3	3
Willowtree School	1	1	4
Medel	2.3	1.8	2.8

*Deltog ej i projektet (Ashtree uppskjutet ett år, Hazeltree implementerade en annan modell).

Två detaljer i tabellen är särskilt värda att notera. Först att de största framstegen tycks ha gjorts i moderna språk, vilket (vi hoppas av en tillfällighet!) var det ämne som hade den minst välbesökta, inledande sessionen. För det andra att det, med undantag för Spruce School, inte verkar finnas någon tendens att TLC- grupper med bäst progression finns i samma skola.

I de skolor som deltog i projektet, var det bara 9 av 21 möjliga TLC- grupper som gjorde rimliga framsteg (definierat som grad 3 eller 4 i tabell 2). En mer noggrann analys av handledarnas loggar visar att den främsta orsaken till det sämre resultatet i de 12 grupperna som gjort långsamma eller inga framsteg (definierade som 2 eller 1 på skalan) var att lärarna inte hade fått avsatt tid att mötas (detsamma var fallet för några av de mer framgångsrika TLC- grupperna, men på grund av sitt intresse och engagemang i projektet, hade de beslutat att ta av sin egen personliga tid till mötena). När mötestid för medlemmar gjorts tillgänglig inom avtalad arbetstid fungerade TLC- grupperna bra, med stor entusiasm för mötena och i synnerhet tycktes lärarna värdera möjligheten att tala om sin praktik på ett strukturerat sätt.

Vid första anblicken är svårigheten för TLC- grupperna att finna tid för att träffas ganska överraskande. Även om ingen av skolorna i Cannington finns med på listan över de cirka 600 skolor i England som utsetts av regeringen som "National Challenge"- skolor, på grund av att de inte lyckats nå riktmärket 30 % av eleverna med godkänd färdighet (grad C eller över) i fem ämnen i den nationella examinationen för 16-åringar, så finns det ändå ett stort tryck på att förbättra elevresultat. Allmänna kommunikationer i Cannington är så bra att elever lätt kan resa från ena sidan av kommunen till den andra. Föräldrarna har ett stort utbud av skolor att välja bland och även om val av skola inte är helt beroende av resultatet för 16-åringar så påverkar det. Samtliga rektorer säger att förbättrade akademiska resultat, mätt med framgång på nationella prov vid 16 och 18 års ålder, är en av deras viktigaste prioriteringar. Och ändå, trots forskning som visar att formativ bedömning kan göra mer skillnad för de akademiska resultaten än något annat, så verkar det vara svårt för rektorer och andra skollädares att prioritera utvecklingen av formativ bedömning.

Det är ännu mer förvånande när man betänker att rektorerna hade gjort ett åtagande om utvecklingsprogrammet sex månader innan det inleddes, hållits fullt informerade om vilken resurs som var nödvändig för de månatliga mötena och hade gjort en betydande investering i projektet genom att avsätta i genomsnitt 12 lärardagar för att lärare skulle kunna delta i introduktionsutbildningen.

Rektorerna inom High School i Cannington står fast vid - och är faktiskt ganska positiva till - projektet. Var och en av dem ser fram emot en "nystart" av *Embedding formative assessment* programmet i hela skolan nästa år; med hjälp medlemmar ur årets grupper som TLC- ledare.

Sammanfattning

Forskning tyder på att när formativa metoder för bedömning integreras minut-för-minut och dag för dag i lärares klassrumsaktiviteter, så sker en betydande ökning av elevers prestationer - i storleksordningen 70 till 80 procents ökad inlärningshastighet - och att det är möjligt att mäta utfallen även med externt godkända standardiserade test. I själva verket tyder tillgängliga belägg på att det inte finns något annat som är till närmelsevis lika prisvärt och som kan förväntas ha lika stor effekt.

Fast det fanns klara belägg för hur lärare kunde ges stöd i att utveckla sin användning av formativ bedömning och det var möjligt att genomföra i sammanhang både i USA och i Storbritannien, så var det mindre klart hur man gör det i större skala. I denna uppsats har vi beskrivit utvecklingen av ett par produkter - *Keeping Learning on Track™* och *Embedding formative assessment* - som verkar vara effektiva sätt att stödja lärare i sin utveckling av formativ bedömning, även när de levereras med lite stöd utöver de resurser som kan få plats på en cd-rom. De två modellerna var avsedda att vara uppskalningsbara och att kunna användas med minimalt tillskott av ytterligare resurser från skolan och inom en tidstilldelning väl inom vad som skolor rutinmässigt spenderar på administration och byråkrati. Vi anser därför att vi har gjort två viktiga steg: att klargöra vad som bör prioriteras för lärares kompetensutveckling och vilken form professionell kompetensutveckling bör ha.

Vad som däremot förvånat oss, är att det tredje steget – att faktiskt få skolorna att prioritera kompetensutveckling - verkar vara svårare än något av de två första. Och det blir en prioritering för vårt framtida arbete.

Referenser

- Allal, L., & Lopez, L. M. (2005). Formative assessment of learning: a review of publications in French. In J. Looney (Ed.), *Formative assessment: improving learning in secondary classrooms* (pp. 241-264). Paris, France: Organisation for Economic Cooperation and Development.
- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, **61**(2), 213-238.
- Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice*, **5**(1), 7-73.
- Black, P. J., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, **21**(1), 5-31.
- Black, P., & Harrison, C. (2002). *Science inside the black box: assessment for learning in the science classroom*. London, UK: King's College London Department of Education and Professional Studies.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Buckingham, UK: Open University Press.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Ed.), *Educational evaluation: new roles, new means: the 68th yearbook of the National Society for the Study of Education (part II)* (Vol. 68(2), pp. 26-50). Chicago, IL: University of Chicago Press.
- Boekaerts, M. (2006). Self-regulation and effort investment. In K. A. Renninger & I. E. Sigel (Eds.), *Handbook of child psychology volume 4: child psychology in practice* (6 ed., pp. 345-377). New York, NY: Wiley.
- Broadfoot, P. M., Daugherty, R., Gardner, J., Gipps, C. V., Harlen, W., James, M., & Stobart, G. (1999). *Assessment for learning: beyond the black box*. Cambridge, UK: University of Cambridge School of Education.
- Brookhart, S. M. (2004). Classroom assessment: tensions and intersections in theory and practice. *Teachers College Record*, **106**(3), 429-458.
- Brookhart, S. M. (2007). Expanding views about formative classroom assessment: a review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: theory into practice* (pp. 43-62). New York, NY: Teachers College Press.
- Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: on procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations In Learning: New Environments for Education* (pp. 291-292). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ciofalo, J., & Leahy, S. (2006). *Personal action plans: helping to adapt and modify techniques*. Paper presented at the Annual meeting of the American Educational Research Association: San Francisco, CA.
- Clymer, J. B., & Wiliam, D. (2006/2007). Improving the way we grade science. *Educational Leadership*, **64**(4), 36-42.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, **58**(4), 438-481.
- Dempster, F. N. (1991). Synthesis of research on reviews and tests. *Educational Leadership*, **48**(7), 71-76.
- Dempster, F. N. (1992). Using tests to promote learning: a neglected classroom resource. *Journal of Research and Development in Education*, **25**(4), 213-217.
- DuFour, R., DuFour, R., Eaker, R., & Many, T. (2005). *Learning by doing: a handbook for professional learning communities at work*. Bloomington, IL: Solution Tree.
- Elshout-Mohr, M. (1994). Feedback in self-instruction. *European Education*, **26**(2), 58-73.
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation - a meta-analysis. *Exceptional children*, **53**(3), 199-208.
- Gawande, A. (2007). *Better: a surgeon's notes on performance*. London, UK: Profile Books.
- Ginsburg, H. P. (2001). *The Mellon Literacy Project: what does it teach us about educational research, practice, and sustainability?* New York, NY: Russell Sage Foundation.
- Gipps, C. V., & Stobart, G. (1997). *Assessment: a teacher's guide to the issues* (3 ed.). London, UK: Hodder and Stoughton.

- Goe, L., & Bridgeman, B. (2006). *Effects of Focus on Standards on academic performance*. Unpublished report. Princeton, NJ: Educational Testing Service.
- Grossman, P., Wineburg, S., & Woolworth, S. (2000). *What makes teacher community different from a gathering of teachers?* Seattle, WA: University of Washington Center for the Study of Teaching and Policy.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, **77**(1), 81-112.
- Hayes, V. P. (2003). *Using pupil self-evaluation within the formative assessment paradigm as a pedagogical tool*. Unpublished EdD thesis, University of London.
- Hodgen, J., & Wiliam, D. (2006). *Mathematics inside the black box: assessment for learning in the mathematics classroom*. London, UK: NFER-Nelson.
- James, M. (1992). *Assessment for learning* Paper presented at the Annual Conference of the Association for Supervision and Curriculum Development (Assembly session on 'Critique of Reforms in Assessment and Testing in Britain') held at New Orleans, LA. Cambridge, UK: University of Cambridge Institute of Education
- Jones, J., & Wiliam, D. (2007). *Modern foreign languages inside the black box: assessment for learning in the modern foreign languages classroom*. London, UK: Granada.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, **119**(2), 254-284.
- Köller, O. (2005). Formative assessment in classrooms: a review of the empirical German literature. In J. Looney (Ed.), *Formative assessment: improving learning in secondary classrooms* (pp. 265-279). Paris, France: Organisation for Economic Cooperation and Development.
- Leahy, S., & Wiliam, D. (2009). *Embedding assessment for learning — a professional development pack*. London, UK: Specialist Schools and Academies Trust. <https://www.schoolsnetwork.org.uk/pages/default.aspx>
- Maher, J., & Wiliam, D. (2007). *Keeping learning on track in new teacher induction* Paper presented at the Symposium entitled "Tight but loose: scaling up teacher professional development in diverse contexts" at the annual conference of the American Educational Research Association held at Chicago, IL.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, **22**(2), 155-175.
- Nyquist, J. B. (2003). *The benefits of reconstruing feedback as a larger system of formative assessment: a meta-analysis*. Unpublished Master of Science, Vanderbilt University.
- Popham, W. J. (2006). Phony formative assessments: buyer beware! *Educational Leadership*, **64**(3), 86-87.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. 1, pp. 39-83). Chicago, IL: Rand McNally.
- Shepard, L. A. (2007). Formative assessment: caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment: shaping teaching and learning* (pp. 279-303). Mahwah, NJ: Lawrence Erlbaum Associates.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, **78**(1), 153-189.
- Stiggins, R. J. (2002). Assessment crisis: the absence of assessment for learning. *Phi Delta Kappan*, **83**(10), 758-765.
- Sutton, R. (1995). *Assessment for learning*. Salford, UK: RS Publications.
- Thompson, M., & Goe, L. (2008). *Models of effective and scalable teacher professional development* (Research report RR-09-07). Princeton, NJ: Educational Testing Service.
- Thompson, M., & Wiliam, D. (2008). Tight but loose: a conceptual framework for scaling up school reforms. In E. C. Wylie (Ed.), *Tight but loose: scaling up teacher professional development in diverse contexts* (RR-08-29, pp. 1-44). Princeton, NJ: Educational Testing Service.
- Wiliam, D. (2005). *Measuring 'intelligence': what can we learn and how can we move forward?* Paper presented at the Annual meeting of the American Educational Research Association held at Montreal, Canada.
- Wiliam, D. (2006). Assessment: learning communities can use it to engineer a bridge connecting teaching and learning. *Journal of Staff Development*, **27**(1), 16-20.
- Wiliam, D. (2007). Keeping learning on track: classroom assessment and the regulation of learning. In F. K. Lester Jr (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1053-1098). Greenwich, CT: Information Age Publishing.

- William, D. (2009). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment*. New York, NY: Taylor & Francis.
- William, D., & Thompson, M. (2007). Integrating assessment with instruction: what will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: shaping teaching and learning* (pp. 53-82). Mahwah, NJ: Lawrence Erlbaum Associates.
- William, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles Policy and Practice*, **11**(1), 49-65.
- Wylie, E. C., Lyon, C. J., & Goe, L. (2009). *Teacher professional development focused on formative assessment: changing teachers, changing schools* (Vol. RR-09-10). Princeton, NJ: Educational Testing Service.
- Wylie, E. C., Lyon, C. J., & Mavronikolas, E. (2008). *Effective and scalable teacher professional development: a report of the formative research and development* (Vol. RR-08-65). Princeton, NJ: Educational Testing Service.
- Yeh, S. S. (2006). *Raising student achievement through rapid assessment and test reform*. New York, NY: Teachers College Press.



Göteborgs Stad
Utbildning