

TEMAPRESENTATION: EXAMINATION OCH BEDÖMNING FÖR LÄRANDE?

Temaredaktör: Stefan Ekecrantz

I det här temanumret behandlas en serie skilda perspektiv på examination, bedömning och styreffekter av olika slag. Den summativa examinationens inverkan på elevers och studenters studiestrategier är ett känt fenomen, även om vissa forskare har börjat ifrågasätta hur allmängiltig dess empiriska bas verkligen är (Jensen, McDaniel, Woodard & Kummer, 2014; Joughin, 2010). En annan aspekt som finns väl beskriven i litteraturen är styreffekter av så kallad *high stakes testing*, när stora, standardiserade kunskapsmätningar används för att kvalitetsgranska skola och högre utbildning (Au, 2007; Dulude, Spillane, & Dumay, 2015). När det gäller formativ bedömning finns på motsvarande sätt en hel del kunskap om dess effekter på lärande – även om det också i det fallet finns anledning att besinna sig vad gäller dess empiriska förankring. Ett fält som däremot är betydligt mindre utforskat är hur ett alltmer framträdande formativt bedömningsparadigm påverkar skolans värld när sådan bedömningsforskning och litteratur anammas av byråkrati och regelverk på alla nivåer.

I sina kvalitetsredovisningar har dagens svenska skolledare och lärarpersonal att anpassa sig till ett sådant bedömningsparadigm och får ibland ros av Skolinspektionen för sina ansträngningar, som när inspektionen lyfter en verksamhet som särskilt förtjänstfull därför att "[s]åväl lärarna som eleverna beskriver också hur Jensen Uppsala jobbar med formativ bedömning" (Skolinspektionen, 2012, s. 3). Ibland får en skola istället ris för motsvarande tillkortakommanden:

Skolinspektionen har i beslut den 7 december 2012 förelagt kommunen att vidta åtgärder avseende området bedömning och betygssättning. [...] Skolinspektionen begär därför följande kompletterande uppgifter: [...] En redovisning av uppföljning och analys av lärares användning av formativ bedömning (Skolinspektionen, 2013a, s. 1).

STEFAN EKECRANTZ

Fil. Dr. i Historia

*Verksam vid institutionen för pedagogik och didaktik
Stockholms universitet, 106 91 Stockholm.*

E-post: stefan.ekecrantz@su.se

Som argument för varför just förekomsten av formativ bedömning är en särskilt angelägen kvalitetsindikator finns regelmässigt hänvisningar till pedagogisk forskning, i alldeles särskild synnerhet John Hatties arbete *Visible Learning* (2009), som oftast presenteras som ensam auktoritet "tack vare sitt omfång och fokus på effekter" (Skolinspektionen, 2010, s. 16). Det är emellertid inte bara förekomsten av, eller bristen på, formativ bedömning som uppmärksammas. Efter en större nationell granskning av undervisningen i SO-ämnena i högstadiet konstaterar tillsynsmyndigheten dessutom att en specifik mängd i en viss kontext var alldeles för lite:

Endast på var fjärde lektion kan Skolinspektionen observera tydliga formativa bedömningar till eleverna. [...] Detta är synd då just formativ bedömning, där eleverna får veta var de befinner sig kunskapsmässigt, vart de ska (kunskapskraven), och vad eleven behöver öva sig i för att komma dit, är en enskild faktor som har visat sig ha en stor påverkan på elevernas kunskapsutveckling (Skolinspektionen, 2013b, s. 23).

Även här utgör Hattie (2009) den huvudsakliga referensen, men den här gången ryms i fotnoten även en hänvisning till Jan Håkanssons och Daniel Sundbergs översikt *Utmärkt undervisning: Framgångsfaktorer i svensk och internationell belysning* (2012). En fråga är då vilket stöd som finns i forskningen för den typen av definitiva föreställningar. Vad hävdar till exempel Håkansson och Sundberg själva om forskningsläget?

Hur effektiv formativ bedömning ser ut varierar alltså, baserat på den kognitiva nivån på det efterfrågade resultatet, som till exempel faktakunskaper, vissa färdigheter eller förståelseinriktade kunskaper, och på omgivande faktorer som nämnts ovan, men även på årskurs och ämne (Håkansson & Sundberg, 2012, s. 217).

En sådan skrivning kan ställas mot föreställningen att formativ bedömning kan tillskrivas en universellt giltigt effekt på lärande – och dessutom en effekt av en viss storlek. Går man till nivån direkt under Hatties syntes, ner till exempelvis Hatties eget arbete tillsammans med Helen Timperley (2007), som är en av de mest citerade metastudierna om feedback, framträder en liknande motbild.¹ Redan där kan man se hårddata som relaterar till de "kognitiva nivåer" som omnämns i citatet av Håkansson och Sundberg ovan. Feedback på okomplicerade uppgifter uppges leda till förbättringar av prestationer med en effektstorlek på 0,55. Men, feedback på komplexa uppgifter ger enligt samma studie ett värde på 0,03, det vill säga har i princip ingen effekt alls (Hattie & Timperley, 2007, s. 85). Av alla listade perspektiv på feedback i Hatties och Timperleys metastudie är det bara den något märkliga kategorin *Task feedback designed to discourage the student* som uppvisar ett sämre resultat (-0,14). Ett annat centralt budskap i den studien är att positiva effekter på lärande förutsätter att återkopplingen ges först när eleven/studenten har försökt förstå någonting och dessutom tror sig ha förstått detta. Vad detta "någonting" representerar i ett SO-ämne i högstadiet ter sig väldigt olika men flertalet av de aktuella kunskapsmålen framstår både som högst komplexa och som kunskapsobjekt som inte alltid kan avgränsas till en enskild lektion.

Specifika föreställningar om att det måste ske en viss mängd observerbar formativ återkoppling i just samhällskunskap, religion och historia i årskurs 7–9 har helt enkelt inte stöd i den forskning som refereras, och inte heller i underliggande nivåer i litteraturen. En mer rättvisande bild av kunskapsläget sammanfattas i Åsa Hirsh och Viveca Lindbergs forskningsöversikt om formativ bedömning som nyligen publicerade för Vetenskapsrådets räkning (2015). Där sammanfattar de kunskapsbehovet och rekommenderar forskning som granskar och problematiserar "den förenklade/instrumentella uppfattningen av formativ bedömning som i viss utsträckning råder" (s. 78). Vidare rekommenderar de att de förenklade slutsatser från forskningen inom området som sprids även måste omfatta reservationer rörande generaliserbarhet, samt att vi saknar tillräcklig kunskap om formativa bedömningspraktikers effekter i skilda ämnen och ålderskategorier.

Samtidigt är det uppenbart att ett beslut eller en rapport från en tillsynsmyndighet inte rimligen kan nagelfaras som pedagogisk grundforskning. Att ifrågasätta förenklade försanthållanden må vara forskningens uppgift men falsifieringsförsök och faktisk falsifiering är inte samma sak. Skolinspektionens, Skolverkets och landets lärarutbildningars utgångspunkter apropå formativ bedömning ter sig trots allt som rimliga, givet forskningsläget i stort. Och, allt sammantaget skulle man kunna hävda att skolans värld idag kanske förhåller sig till aktuell forskning i större utsträckning än många andra jämförbara samhällssektorer, även om förstås många önskar mer av den varan. Ändå gör man någonting speciellt när man motiverar sina utgångspunkter och ställningstaganden med just hänvisningar till stora volymer av empirisk forskning och därmed frammanar en bild som kan uppfattas som definitiv. Måste till exempel en skola som underpresterar nödvändigtvis öka mängden formativ bedömning eller inte, och hur kommer en implementering av ett sådant påbud landa i sin tillämpning? I en nyligen publicerad studie om en stor bedömningsreform i Borås kommun visar Anders Jönsson, Christian Lundahl och Anders Holmgren (2015) bland annat hur lärarna trots fleråriga utbildningsinsatser i praktiken kom att uttolka konceptet bedömning för lärande främst som egen, arbetsintensiv återkoppling till eleverna – i strid med den betydligt mer sammanstatta bilden som ges i litteraturen.

Hur forskning som har filtrerats genom lager av förenklingsprocesser landar i skolans värld framstår därför som ett angeläget forskningsområde i sig. Är det bara förenklingar av förenklingar av förenklingar, eller händer det någonting kvalitativt nytt på vägen? En brobyggare kan förlita sig på Newtons 300 år gamla mekanik och är mindre betjänt av Einsteins och Bohrs tillkrånglingar av densamma. Är det den typen av förenklingar som det handlar om? I viss mån ja. Skolledare och andra praktiker har begränsad nytta av många av de inomvetenskapliga perspektiv som skärskådas i pedagogisk forskning, men vital information riskerar också att försvinna i processen. När medicinsk forskning ska tillämpas av vårdpersonal är både rekommenderad behandling och graden av osäkerhet helt avgörande information. En viss medicinering och dos kan utgöra en evidensbaserad utgångspunkt, men om det är känt från forskningen att det kan skilja sig väldigt mycket åt på individnivå så krävs uppföljning och kalibrering. Reservationer om osäkerhet är

i det exemplet inte detaljer som kan silas bort på vägen utan är helt avgörande delar av huvudresultatet som måste följa med ända fram till brukarledet. Pedagogisk grundforskning må ofta producera frustrerande det-beror-på-resultat, men man kan också hävda att många reservationer om osäkerhet och kontextberoende på precis samma sätt som i medicinsk forskning behöver betraktas som del av huvudresultatet, och kan inte alltid avfärdas som detaljträd som skymmer skogen.

Som de inledande exemplen ovan visar kan John Hatties genomslag knappast överdrivas. Det som gör hans syntes unik är dels själva storleken, dels att han för samman och jämför metastudier från skilda fält. Det gör att skilda pedagogiska interventioners effektstorlekar kan ställas mot varandra. *Pre-term birth weight* är bra (0,54), *student control over learning* är meningslöst (0,04) och *summer vacation, welfare policies* och *mobility* är skadligt (-0,09; -0,12 och -0,34). Eller...? När det gäller enskilda områden, som formativ bedömning och återkoppling, ter det sig dock mer relevant att luta sig mot specifik litteratur inom just det fältet. I det fallet är Paul Blacks och Dylan Wiliams arbeten från sent 1990-tal fram till idag mer naturliga referenser – i den mån det över huvud taget är lämpligt att förlita sig på enskilda auktoriteter.

Ett möjligt skäl till genomslaget för alla dessa tre ligger just i den metastudiegenre som de har verkat i. Denna har per definition långtgående generaliseringsanspråk och utmynnar dessutom i en (förment) lättförstådd kvantifiering av effekter. Under senare tid har dock effektstorlekar som utbildningsvetenskapens *lingua franca* börjat ifrågasättas alltmer, dels i form av ifrågasättande av metodologin i sig, dels av hur den har kommit att uttolkas. I en nätdiskussion om bland annat effektstorlekar skriver Dylan Wiliams själv om sina dubier idag och om möjliga tillkortakommanden i hans och Paul Blacks tidigare arbeten:

To be honest, however, while we realized there were some problems with using effect sizes (one section of the academic paper is entitled "No meta-analysis"), it is only within the last few years that I have become aware of just how many problems there are. Many published studies on feedback, for example, are conducted by psychology professors, on their own students, in experimental sessions that last a single day. The generalizability of such studies to school classrooms is highly questionable. Another point that I have only recently understood well is the impact of the limited power of most educational and psychological experiments on meta-analyses. [...] In retrospect, therefore, it may well have been a mistake to use effect sizes in our booklet "Inside the black box" to indicate the sorts of impact that formative assessment might have. (Wiliam, 2014).

En liknande problematik utvecklar sig Stefan Ekecrantz om i den inledande artikeln i det här temanumret, "Feedback and student learning? – A critical review of research". Via vad som benämns som *a genealogical case study* (en "släktforskande fallstudie") följer författaren mer generella utsagor om feedback och lärande ner till en särskilt inflytelserik metastudie av Avraham Kluger och Angelo DeNisi (1996). Den studien bygger i sin tur på 131 empiriska feedbackstudier som vid en närmare analys visar sig inte alls stödja de påståenden som efterföljande forskning ofta har påstått. En samlad beskrivning är att den underliggande forskningen knappast alls uttalar sig om feedback och elevers/studenters lärande, utan istället

mestadels handlar om beteendemodifikation och uthållighet i olika miljöer. Fallstudien är tänkt att fungera som ett av många möjliga exempel på hur processen mellan empirisk grundforskning och sammanfattande synteser kan te sig.

I artikeln "Alone with the test – Students' perspectives on an enacted policy of national testing in Swedish schools" undersöker Håkan Löfgren och Ragnhild Löfgren hur elever i årskurs 6 erfar nationella prov i NO- och SO-ämnena. Genom en serie gruppintervjuer visar författarna hur olika dessa prov upplevs, om olika former av press från hemmet och om högst fragmentariska förberedelser i skolan. Ett huvudresultat i studien är att eleverna i huvudsak upplever sig ensamma med uppgiften och att de tvingas söka sig till andra elever för att försöka förstå hur de ska närma sig den. Resultaten relateras bland annat till policy- och implementeringsteori, där elevernas identitetsformering tolkas som effekter av lokala praktikers *enactment* av policyn.

Per-Åke Rosvall tar i artikeln "Examinationer för elevinflytande?" upp flera eftersatta perspektiv på den summativa examinationens möjliga styreffekter. I en etnografisk studie av en gymnasieklass har författaren via klassrumsobservationer, insamlat skolmaterial och intervjuer med elever, lärare och rektor studerat bland annat relationen mellan examinationspraktik och elevinflytande. I en jämförelse mellan två lärares tämligen skilda arbetssätt och skilda sätt att uppmuntra till aktivt inflytande på undervisningens innehåll och form, visade det sig att eleverna i båda fallen i förstone förhöll sig till den befintliga examinationen. Något anmärkningsvärt motsatte sig till och med eleverna ett eget inflytande över undervisningen. Med stöd i studien argumenterar Rosvall bland annat för att åtgärder för ökat elevinflytande kan misslyckas om inte ett sådant inflytande även omfattar examinationens form och innehåll. Vidare diskuteras relationen mellan examinationens utformning och elevernas lärande.

Temanumret avslutas med en artikel av Jens Gardesten och Henrik Hegender, "Underkännanden inom verksamhetsförlagd lärarutbildning. Resultat från en forskningsexpedition i svårframkomlig terräng". Studien fokuserar på organisatoriska och administrativa perspektiv och utgör ett första delprojekt inom ramen för ett större projekt om underkännanden i svensk lärarutbildning. I ett första steg samlades enkätdata in från 29 VFU-koordinatorer vid 20 svenska lärosäten. Med utgångspunkt i denna inhämtades sedan kompletterande uppgifter via telefon- och e-post från samtliga 25 lärosäten med lärarutbildning. Undersökningen visar att andelen underkända studenter kunde variera kraftigt, med ytterlighetsfallen noll respektive nio procent, med medelvärden på någon enstaka procent ($M=2,9$; $Md=1$). I studien kartläggs vidare bland annat vanligt förekommande processer när ett underkännande är tycks vara förestående respektive arbetssätt och rutiner efter underkännanden. I en avslutande diskussion redogörs även för några preliminära resultat från det större projektet, inklusive jämförelser med andra utbildningar med verksamhetsförlagd utbildning.

NOTER

1. En kritik av bland annat just Hattie och Timperleys arbete presenteras i artikeln "Feedback and student learning? – A critical review of research" i det här temanumret. I just den här diskussionen problematiseras emellertid inte deras utgångspunkter då argumentationen relaterar direkt till Skolinspektionens hänvisning.

REFERENSER

- Au, W. (2007) High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Dulude, E., Spillane, J.P. & Dumay, X. (2015) High stakes policy and mandated curriculum: A rhetorical argumentation analysis to explore the social processes that shape school leaders' and teachers' strategic responses. *Educational Policy*, (In press).
- Hausknecht, J.P., Halpert, J.A., Di Paolo, N.T. & Moriarty Gerrard, M.O. (2007) Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92(2), 373-385.
- Hirsh, Å. & Lindberg, V. (2015) *Formativ bedömning på 2000-talet: En översikt av svensk och internationell forskning i Forskning och skola i samverkan*. Stockholm: Vetenskapsrådet.
- Håkansson, J. & Sundberg, D. (2012) *Utmärkt undervisning: Framgångsfaktorer i svensk och internationell belysning*. Stockholm: Natur & Kultur.
- Jensen, J., McDaniel, M., Woodard, S. & Kummer, T. (2014) Teaching to the test... or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, 26(2), 307-329.
- Jonsson, A., Lundahl, C. & Holmgren, A. (2015) Evaluating a large-scale implementation of Assessment for Learning in Sweden. *Assessment in Education: Principles, Policy & Practice*, 22(1), 104-121.
- Joughin, G. (2010) The hidden curriculum revisited: A critical review of research into the influence of summative assessment on learning. *Assessment & Evaluation in Higher Education*, 35(3), 335-345.
- Kluger, A. & 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.
- Skolinspektionen (2010) *Framgång i undervisningen: En sammanställning av forskningsresultat som stöd för granskning på vetenskaplig grund i skolan*. Skolinspektionen, Dnr 2010:1284, 1-18.
- Skolinspektionen (2012) *Beslut för gymnasieskola efter riktad tillsyn av Jensen Uppsala i Uppsala kommun, 2012-11-23, Dnr 400-2011:6483*.
- Skolinspektionen (2013a) *Protokoll efter riktad tillsyn av bedömning och betygssättning i Kumla skola i Tyresö kommun, 2013-08-14 Dnr 430-2011:6483*.
- Skolinspektionen (2013b) *Undervisning i SO-ämnena år 7-9: Mycket kunskap men för lite kritiskt kunskapande*. Skolinspektionens rapport 2013:04. 1-38.
- William, D. (2014) Orubricerad bloggkommentar, 25 januari, 2014. Hämtad 31 augusti, 2015, från <http://www.learningspy.co.uk/myths/things-know-effect-sizes/>