

Att legitimera nationella prov i Sverige och i Norge – en fråga om profession och tradition

CHRISTIAN LUNDAHL & SVERRE TVEIT

Institutionen för humaniora, utbildnings- och samhällsvetenskap
Örebro universitet

Institutt for pedagogikk, Universitetet i Oslo

Artikeln utforskar utvecklingen av nationella bedömningsinstrument, så som nationella prov, i svensk och norsk grund- och gymnasieskola. Fokus riktas mot hur två annars jämförbara länder har hanterat reformering av sina bedömningssystem på olika sätt. Nationella bedömningsinstrument utvecklas och förändras över tid och artikeln knyter dessa förändringar av funktion och syfte till konflikter mellan aktörer så som ämbetsmän, forskare och lärare. Frågan om legitimitet är här helt central. Vi identifierar hur mätbarhet och användbarhet blir ifrågasatt i professionskamper om betyg, prov och bedömning. Historiskt sett har en ökad tonvikt på att använda data från skolans bedömningar en "output"-orienterad styrning av utbildningssystemet ofta lett till att de nationella bedömningsinstrumentens legitimitet utmanas. Bedömningsinstrumentens sviktande legitimitet kan ta sig olika uttryck från tid till annan, och hanteras av staten på olika sätt. Vi menar det nuvarande intresset för formativ bedömning, både hos svenska och norska myndigheter, delvis bottnar i ett behov att få legitimitet också för statens mer summativa bedömningsinstrument.

INLEDNING

Idéhistorikern Karin Johannisson har ur ett historiskt perspektiv beskrivit hur viktiga läkarens bedömningar, eller diagnoser, har varit för läkarkårens status: "För läkaren bekräftar diagnosen hans/hennes professionella kompetens och förmåga att spåra, tolka och klassificera" (2006, s. 30). Samma typ av resonemang har Bourdieu och Passeron (1970/1990) lyft i en tolkning av det franska examensväsendets historia. Lärarkårens legitimitet och självbestämmande bottnar i hur omvärlden ser på skolinstitutionens bedömningar. Bourdieu och Passeron (1970/1990) menar att det inte räcker att förstå kunskapsbedömningar ur ett samhällsperspektiv, det vill säga utifrån samhällets behov av sortering. Man behöver också ställa frågor om vad examinationer fyller för funktion inom ramen för skolinstitutionens autonomi och dess legitimitet, samt därigenom för lärares professionalisering.

I Frankrike, menade Bourdieu och Passeron, kunde examinationer motiveras utan att de knöts till samhällets behov av utbildad arbetskraft (*ibid*). Det franska utbildningssystemet hade lyckats förvärva en relativt stark autonomi i förhållande till andra system i samhället, och examinationerna var en av anledningarna till att utbildningssystemet lyckats erövra denna autonomi och upprätthålla den. Det var genom de examinationer som skolväsendet hade exklusivitet på, som det bestämdes vilka kunskaper som var värdefulla, samt i vilken mån individen förvärvat dem. Genom denna autonoma position kunde utbildningssystemet sörja över återväxten inom den egna lärarkåren genom att sätta upp vissa normer för vilka kunskaper en lärare behöver. Men för att också lyckas upprätthålla autonomi i ett mer komplext samhälle måste skolsystemet lyckas översätta nya yttre förväntningar, som att utbildning ska kunna leda till andra yrken än läraryrket, till sina egna normer för kunskap och kunskapsförvärvande. När utbildningssystemets sätt att klassificera individer efter de kunskaper som erbjuds inom detta system också blir grunden för samhällelig ordning, erhåller utbildningssystemet sin samhälleliga legitimitet. En strategi det franska utbildningssystemet använt för att övertyga samhället om att just dess bedömningar av individens potential var mest relevanta, var genom att hävda att elevernas individuella resultat uppnått i hård tävlan med andra (Bourdieu & Passeron (1970/1990).

Legitimitet betyder att något är berättigat, rättfärdigt eller accepterat. Att legitimera betyder att rättfärdiggöra, exempelvis varför man ska ha nationella prov. Inom organisationsteori, där legitimitet är ett centralt begrepp, skiljer man mellan yttre och inre legitimitet, beroende på om legitimitet ska uppnås innanför eller utanför en organisation (Andersen, 2009). Vi riktar uppmärksamheten mot hur nationella bedömningsinstrument får legitimitet inom lärarprofessionen men också mot den legitimitet lärare har som bedömare sett från samhällets sida. Vad händer när samhället inte längre litar på

lärares bedömningar? Hur går det då att ge dem mer legitimitet? En variant är att förse lärares bedömningar med ökad objektivitet, till exempel genom att basera dem på vetenskapligt utformade tester eller nationella prov; en annan är att lärare visar ökad kompetens och kunnande om bedömning eller att man använder särskilt kompetenta externa examinatorer.

I artikeln tar vi vår utgångspunkt i skillnaden mellan amerikansk och europeisk bedömningstradition och pekar på hur Sverige inkorporerat ett amerikanskt bedömningssynsätt, medan Norge mer upprätthållit en europeisk tradition. Syftet är att förstå hur två olika typer av bedömningstraditioner kan växa fram och få legitimitet där den ena utmanar lärarkårens autonomi och professionalitet i bedömningsfrågor medan den andra bejakar den.

Vi beskriver hur nationella bedömningar, främst genom centralt konstruerade prov, varit stora reformer i sig som behövt legitimeras, men också använts i andra reformer för att ge dessa legitimitet. Vi menar här att Sverige både legitimerat centrala prov med anspråk på att de utformas vetenskapligt och att dessa prov sedan kan legitimera andra reformer i kraft av objektiva resultatmått. Norge däremot har av olika skäl i mindre grad kunnat legitimera centrala prov med hänvisning till vetenskap utan söker hela tiden stöd för dem i lärartraditionen. Den slutsats vi drar är att skolans bedömningar är utsatta för olika professionsstrider där olika anspråk på mätbarhet och användbarhet står i centrum. När lärares autonomi i dessa frågor urvattnas riskerar också lärarnas status att sjunka, samtidigt som de centrala provens status kan öka om de just ger sken av objektivitet. Detta fungerar dock bara till en viss gräns – nationella bedömningsinstrument tappar i förtroende om de står allt för mycket i vägen för lärandet eller lärarna. Vi hävdar att dagens fokus på formativa bedömningar är ett uttryck för att lärarkåren inte accepterar för stark utifrån kommande styrning och nu försöker återerövra förlorad mark som kunskapsbedömande profession.

TIDIGARE FORSKNING

Tidigare historisk forskning om bedömning i de skandinaviska länderna har visat att bedömningar av olika slag under 1900-talet knöts allt starkare till psykologisk och pedagogisk vetenskap (Lundahl, 2006; Ydésen, 2011). Det har till stor del att göra med att bedömningar i skolan fått allt större betydelse för allt fler ifråga om sortering, urval och tillträde. Dessutom har bedömningar i kraft av att vara vetenskapliga, genom den legitimitet vetenskap i största allmänhet kan skänka, använts i politiska beslut och reformer (jämför Lundahl & Waldow, 2009). Det finns emellertid vissa skillnader mellan de skandinaviska länderna där särskilt Norge skiljer sig från Danmark och Sverige i det att olika vetenskapliga kunskaps- och färdighetstest inte används

lika mycket inom skolan, och att det inte funnits en stark psykometrisk forskningstradition så som i Sverige och Danmark (Ludvigsen, Lundahl & Ydésen, 2013).

Ideologiska maktkamper kan delvis förklara bristen på forskning om bedömning i Norge. År 2004, strax före skolreformen "Kunnskapsløftet", infördes nationella prov. Sedan dess har det skett en ökad satsning på kartläggningsprov [kartleggingsprøver] och användning av resultatdata från nationella prov, lärarnas betygssättning och examensprov i styrningen av skolväsendet (Stortingsmelding nr 16, 2006-2007). De senaste åren har det riktats stor kraft på att optimera de nationella proven som del av ett ramverk för kvalitetsbedömning och -utveckling (Tveit, 2014b). Dessa prov är färdighetsprov som inte används för att bedöma elevers kunskapsnivå som del av betygssättningen. Fokus i den här artikeln är på de nationella instrument som används för kunskapsbedömningar i ämnen/kurser (det vill säga den norska centralt givna examen och svenska nationella prov).

En överväldigande del av den svenska forskning som producerats kring bedömningar har varit relaterad till de statligt beställda proven (Lindberg, 2005). Sedan 1940-talet har den varit mer eller mindre sammanflätad med svenska skolmyndigheter och har betraktat sig själv som R&D forskning (research and development) (Lindblad, 1991; Lundahl, 2009). Den betydelse centralt utarbetade prov av olika slag haft under ett halvt sekel har också bidragit till en obalans i forskningen om och med bedömningar. Exempelvis har det funnits brist på mer övergripande analyser av bedömning som historiskt och kulturellt fenomen, bedömningarnas konsekvenser för lärare och elever, eller studier av hur bedömningar används i utbildningspolitik och skolförvaltning. På senare år har det dock kommit en hel del från skolmyndigheterna fristående forskare vilka intresserar sig för bedömning såväl ur ett utvecklingsperspektiv som ur ett kritiskt perspektiv (för en översikt, se Lindberg & Forsberg, 2010). Relationen mellan bedömning, legitimitet och lärares profession är däremot ett relativt svagt utforskat område (jämför Lundahl, 2006, 2009; Mickwitz, 2011). Samtidigt finns en hel del professionsforskning kring läraryrket vilken har betydelse för hur vi tänker kring hur centrala provs legitimering hör ihop med lärares autonomi och professionalitet.

Professionsforskare brukar med begreppet professioner avse speciella arbeten som kräver lång specialiserad utbildning som vilar på vetenskaplig grund. Med specialiserad utbildning avses som regel universitets- och högskoleutbildning eller fackskolor. Kunskaperna testas och leder till examina. Ett examensbevis gör att professionerna kan åberopa monopol på sin tjäns-

temarknad. Vidare kräver de professionella autonomi och vill ha inflytande över sin egen yrkeskultur och sina egna etiska regler och ser ogärna att lekmän försöker kontrollera yrket. Den kontrollen vill man själv utöva (Persson, 2008; Florin, 2010).

Flera forskare pekar dock på hur denna typ av professionalitet inom skolektorn varit svår att upprätthålla, man talar om en de- eller avprofessionalisering. Ingrid Carlgren menar exempelvis att forskning och utbildningspolitiska direktiv vid enhetsskolans tillblivelse kom att orienteras mot gripbara metoder och pedagogiskt material för undervisningen. Denna hållning lade grunden till en objektivisering av lärarna i den svenska pedagogiska forskningen och i utbildningspolitik. Staten började tänka åt lärarkåren (Carlgren, 1986, 1999). Andra har pekat på hur den serie av externa styrmedel som introducerades i samband med decentraliseringen i Sverige under 1990-talet ytterligare minskade lärares handlingsutrymme och status (se vidare Florin, 2010). I Storbritannien upplever lärare att fokus mindre och mindre hamnar på validitet när mängden externa test ökar, och att detta påverkar lärarnas undervisning och egna bedömningar (Black, Harrison, Hodgen, Marshall & Serret, 2010). Andra studier visar att lärare gärna använder bedömningsstöd och att lagom mycket central inblandning har högre legitimitet bland lärarna en helt centralt konstruerade test (Collins, Reiss & Stobart, 2010).

Åtgärder har inte saknats för att återge lärarkåren professionell status. I den processen har klassiska professionella strategier ingått: forskning utifrån lärarnas behov, en lärarauktorisering genom yrkeslegitimation, formulering av etiska regler et cetera. Men det verkar som de gamla strategierna inte biter, skriver Florin (2010). Genom att jämföra svensk bedömningstradition med norsk, vill vi se under vilka premisser det går att kombinera uppkommande krav och förväntningar på skolan med läraryrkets behov av kontinuitet och autonomi.

Vad gäller jämförelser mellan Sverige och Norge har det gjorts många statsvetenskapliga och skolhistoriska komparativa analyser (bland annat Sejersted, 2005; Oftedal Telhaug & Mediås, 2005). Komparativ forskning har jämfört intelligensmätning (Ludvigsen, Lundahl & Ydesen, 2013), men det finns inte särskilt många systematiska jämförelser om bedömning i skolan mellan Norge och Sverige. Snarare används grannlandet som en utblick i rapporteringar innanför det enskilda landets kontext, (för Norge: Lysne, 1999, 2004; i Sverige: Lundahl, 2009, 2011). Det är därför särskilt intressant att försöka förstå hur två så närliggande länder kan ha utvecklat så pass olika bedömningstraditioner, så som vi hittills antytt.

METOD

Den här artikeln bygger på analyser och jämförelser av material vi tidigare samlat in och sammanfattat var för sig (Lundahl, 2006, 2009; Tveit, 2014a, 2014b). Dessa arbeten bygger i sin tur på flera olika källmaterial bestående av vetenskapliga artiklar, lärartidningar, dagspress, offentliga utredningar, myndighetsrapporter, websidor och på muntliga källor. Inom pågående projekt (Tveit, pågående) har vi genomfört intervjuer med nyckelpersoner runt provverksamheten i Sverige (Skolverket) och Norge (Utdanningsdirektoratet). I vår historiska genomgång av svensk och norsk bedömningstradition har vi utgått från diskussionen inom komparativa studier om så kallade funktionella ekvivalenter (Schriewer, 2012), på så sätt att vi försöker identifiera likheter och skillnader mellan fenomen som till synes är ekvivalenta. Hur skiljde sig till exempel de normalfördelade proven åt mellan Sverige och Norge? Svaret på den frågan kan utmana nationsspecifika uppfattningar exempelvis om legitimitet. Genom att anlägga en jämförande blick, såväl historiskt som mellan länder, och med fokus på legitimitet kan vi blottlägga en del av de inbyggda spänningar och dilemman som finns med bedömningar och hur de används i olika syften.

Både i Sverige och Norge utgör centralt konstruerade prov en väsentlig del av betygsättningen i grundskolan och i gymnasiet, men de svenska nationella proven och de norska examina har olika funktioner. Beskrivningen av det norska examinationssystemet avgränsas här till de centralt utarbetade skriftliga examina, som vi benämner "nationella examina". De svenska nationella proven betygsätts av lärarna själva, medan de norska nationella examina betygsätts av två externa examinatorer – särskilt kompetenta lärare – som enas om ett betyg. Detta betyg hamnar sedan på elevens betygskort jämte det betyget elevens lärare satt, medan eleverna i Sverige "bara" får ett betyg för varje ämne.

BEDÖMNINGSTRADITIONER

Utifrån John Carsons (2006) utläggning om de franska och amerikanska meritokratiska traditionerna går det att identifiera en europeisk respektive amerikansk tradition i de skandinaviska ländernas betygssättningspolitik. Den norska traditionen är starkast påverkad av Kontinentaleuropas tonvikt på expertbedömningar, manifesterat genom examenssystemet. Sverige, däremot, är antagligen det europeiska land (möjligen jämte England och Skottland) som i störst grad har påverkats av den amerikanska utvecklingen av psykometriska tester som fick sitt genombrott under och efter första världskriget (Lawn, 2008). Ytterst handlar dessa olika traditioner om, menar Carson (2006), olika sätt att legitimera social stratifiering baserad på utbildningsmeriter.

Den europeiska och den amerikanska meritokratiska traditionen

Med den sociala rörelsen under och efter den franska revolutionen (1789-1799) växte nya ideal fram för social selektion. I Frankrike framhölls att grunden för meritokratiska beslut skulle vara *medfödd* talang demonstrerad i konkurrens som var öppen för alla. Det lades vikt på formell utbildning och kompetens som skulle dokumenteras genom examinationer. I Amerika växte det istället fram en uppfattning om att varje enskild individ kunde uttrycka sin talang och uppnå framgång. Amerikanerna la därför mer vikt på personliga egenskaper än på formell utbildning som ett medel för social selektion. På det sättet växte mätning av individers personliga egenskaper fram som det viktigaste underlaget för urval. Senare har man kunnat slå fast att dessa tidiga mätningar i hög grad var kulturellt betingade där social bakgrund fick stort genomslag. Till första världskriget utvecklades dessa idéer som byggde på individmätningar till grupptest att använda vid mönstringen, de så kallade *army alpha* och *army beta* testen (det senare för illitterata). Den amerikanska militären använde intelligenstest i stor skala under första världskriget och nästan två miljoner soldater testades och sorterades (Carson, 2006). De första intelligenstesten som utvecklades av Alfred Binet och Theodore Simone hade varit individbaserade. Efter första och andra världskriget fick gruppbaseade intelligenstest ett enormt genomslag, inte minst i utbildningssektorn, där de kunde användas för jämförelser och gallringar.

De ursprungliga svenska och norska skolexaminationerna

Såväl Sverige, Danmark och Norge har långa traditioner av att selektion till vidare utbildning baserades på examinationer genomförda av lärare, rektorer och censorer. Under 1800-talet utvecklades läroverken i städerna för att tillfredsställa den nya växande medelklassens behov. Det var genom examensväsendet som värdet av gymnasial utbildning skulle komma att regleras, och vi ser under senare delen av 1800-talet ett flertal olika reformer av examen. Examen blir ett sätt att styra utbildningens innehåll och säkerställa undervisningens kvalitet.

I Sverige bestämdes i 1693 års skolordning att lärjungarna skulle examineras både vid utträdet från en skola eller utbildningsnivå och vid inträdet till nästa. Det framgår också att lärare kunde förlora sitt ämbete om en överordnad lärare ansåg att han givit felaktiga vittnesmål, det vill säga omdömen, om eleven. Denna ”dubbla examination” var praktik i Sverige fram till 1862, då det beslutades att man skulle kombinera gymnasielärares bedömningar med *censorer* från universiteten (Lundahl, 2010).

Enligt den nya examensstadgan 1862 skulle de avlämnande lärarna själva sätta vitsord på lärjungarna baserat på skriftliga provning. Censorerna

medverkade sedan vid muntliga förhör av studenterna i ett slags övervakade funktion. Om censorernas mening skiljde sig från lärarnas hade censorerna utslagsröst. I gymnasieskolan skulle också medbedömare utses, och innan censorerna kom in i bedömningen hade således först kollegiet i enighet, eller genom två-tredjedels majoritet, beslutat om skolans vitsord över eleven. På så vis uppstod en slags interbedömarkontroll i två steg där censorerna hade störst inflytande. I ett system med en absolut kunskapssyn var detta rimligt och censorerna kan betraktas som förvaltare och överförare av vad som skulle bedömas som viktiga kunskaper. Systemet var en form av hierarkisk *intern kontroll*. Vissa mindre justeringar av denna, annars fram till avskaffandet på 1960-talet oförändrade institution, genomfördes till gymnasielärarnas fördel i den meningen att de fick större autonomi, självbestämmande, över examinationen (Sjöstedt & Thurén, 1963; Lundahl, 2010).

I den svenska folkskolan dröjde det till införandet av normalplaner (kursplaner) 1878 och uppstramningen av skolplikten innan man fick de motiv till uppflyttningsförhör och slutbetyg som länge funnits inom läroverken. De förhör som hölls i folkskolan hade främst varit knutna till läxläsningen och betraktades som en pedagogisk metod (jämför Lundahl, 2006). Slutförhör syftande till avgångsexamen föreskrevs från 1877, och att avgångsbetyget skulle ges för insikter, färdigheter, flit och uppförande reglerades 1882, men graderande betyg och bestämda betygsbeteckningar över kunskaper, färdigheter, flit och uppförande kom först i och med 1897 års stadga för folkundervisningen (standardiserade betygssteg hade införts i läroverken redan 1820). De beteckningar som infördes: ”berömlig, med utmärkt beröm godkänd, med beröm godkänd, med nöje godkänd, godkänd, icke fullt ut godkänd, underkänd” (Kongl. Maj:ts Förnyade Nådiga Stadga angående Folkundervisningen i Riket 1897 § 47. s. 46) hade någorlunda samma utformning fram till den relativa 1 – 5 skalans införande 1962 (jämför SOU 1942:11, bilaga M; Andersson, 1999; Skolverket, 2014).

Norge var i union med Danmark och styrdes av danska lagar fram till 1814. I Danmark-Norge, precis som i Sverige, ställdes skolan till svars för om en elev som inte var mogen för uppgiften skickades till universitetet. År 1775 fick man bestämmelser om att det skulle utfärdas böter på 20 riksdaler om någon av skolans inskrivna kandidater misslyckades på ”examen artium”, vilket sedan förstärktes med att rektorn framtogs ämbetet om en kandidat underkänts tre gånger. Med *Forordning angaaende de lærde skoler* för Danmark-Norge från 1809 fick man för första gången ett formellt reglemente för gymnasielärarnas betygsättning. Universitetet i Köpenhamn och senare i Oslo (etablerat 1811) kontrollerade emellertid fortfarande examen artium, som fungerade som ett antagningsprov till högre utbildning. På den tiden bestod examen artium av åtta skriftliga prov, fördelat på tre dagar (Lysne, 1999).

Byskoleloven og landskoleloven från 1848 slog fast att den offentliga examinationen skulle övervakas av en präst (Lysne, 1999:68). *Lov om Almueskolevæsenet paa Landet* från 1860, förberedd av Hartvig Nissen, fastslog att prästen inte längre skulle "holde", men "overvære" examen. Nya regleringar för "middelskoleeksamen" implementerades 1873, men detta hade inga konsekvenser för examen artium. (Lysne, 1999). En mer genomgripande förändring kom 1884, när Johan Sverdrup blev statsråd och inledde arbetet med enhetsskolan, och examen artium överfördes från universitetet till godkända gymnasieskolor (Lysne, 1999). Gymnasielärarna hade då, sedan 1859, varit accepterade som censorer vid examen artium. När gymnasielärarprofessionen 1884 fick full kontroll över examen artium, var det ett resultat av en lång kampanj för ökat erkännande av deras ämnesmässiga kompetens att veta vilka kunskaper som behövdes på universitetet (Lysne, 1999). Vi kan alltså se att skolan både i Sverige och Norge under 1800-talet präglades av ett examenssystem främst inriktat på att tillgodose kunskapsyrkenas egna reproduktion, i linje med hur Bourdieu och Passeron (1979/1990) beskrivit det franska systemet under samma period. I takt med utbildningsväsendets ökade samhällsnytta kom dock skolinstitutionens egna system för bedömning att utmanas.

Provens roll i efterkrigstidens utbildningsreformer

Både i Sverige och i Norge intensifierades planerna på en enhetsskola efter andra världskriget och i båda länderna satsades i samband med det på pedagogisk forskning. En viktig fråga handlade om att kalibrera kursplaner så att de var väl avstämda med vad det "normala barnet" kunde tänks klara av. Här väljer dock Sverige och Norge lite olika tillvägagångssätt. I Sverige hade det sedan slutet av 1800-talet funnits en strävan mot att göra folkskolan till en gemensam bottenkola för alla och därmed avskaffa det så kallade parallellskolesystemet där samma årskurser kunde läsas i olika skolformer (där realskolan var mer studieförberedande och hade högre status än folkskolan). Besluten under 1940-talet att verka för en enhetsskola ledde till att en rad utredningar och utvärderingar sjösattes.

Norge använde Sverige i stor utsträckning som en modell men hade bättre förutsättningar för att etablera en enhetsskola eftersom man inte hade ett parallellskolesystem. Norge hade varit ett fattigt land, och försiktiga försök att göra om skolsystemet till ett parallellskolsystem hade av bland annat ekonomiska skäl inte fått genomslag. Därför följde man det svenska reformarbetet med stort intresse. Ledande ämnespedagoger, skolinspektörer, lektorer och lärare "myllrade" över gränsen till grannlandet för att lära av svenskarna (Telhaug, 2003).

I Norge hade man emellertid inte samma tradition för centrala byråkratiska organ som i Sverige. Efter modell från den svenska *Skolöverstyrelsen*

upprättades *Forsøksrådet for skoleverket* 1954 (Sejersted, 2005). Precis som i Sverige tog det några år, men genom en utvidgad försöksverksamhet fick man efterhand nioårig grundskola 1962 (Tønnessen, 2011). Detta gav nya möjligheter och förutsättningar för prov och betygsättning i båda länderna. Kursplaner skulle revideras och gränser för skolmognad och hjälpklasser dras. Enhetsskolan förde med sig behov av nya kunskaper, vilka i Sverige till stor del kom att bygga på prov och tester av olika slag (Lundahl, 2006). Både tillskapandet av en enhetsskola och ”driften” av den förutsatte med andra ord nya standards (det vill säga ”rättesnören”). I Sverige kom dessa att utgå allt mer från matematiska normer, medan Norge förlitade sig mer på etablerade praktiker.

Svensk och norsk bedömningstradition skiljer sig åt under efterkrigstiden

Både det svenska och norska reformarbetet kring prov och bedömning hade sitt ursprung i 1930-talet, då man deltog i ett stort internationellt utvecklingsarbete – *The International Examination Inquiry* – lett av amerikanska forskare (se vidare Lawn, 2008). Särskilt svenska pedagoger hade täta relationer med amerikanska forskare, och blev i efterkrigstidens Europa pionjärer när man la tonvikt på utvärdering som grund för förbättring av resultaten i offentliga tjänster generellt (Lundahl & Pettersson, 2010). Som en av de svenska forskarna bakom *The International Examination Inquiry*, identifierade Frits Wigforss i flera studier under 1930 och 1940-talet svaga korrelationer mellan svenska elevers prestationer på antagningsprov och senare prestationer i realskolan (Lundahl, 2008, 2009). Inträdesproven till realskolan hade med andra ord låg predikativ kraft. Eftersom de också ansågs styra arbetet i folkskolan på ett negativt sett, menade Wigforss att det var bättre att använda betyg som inträdesgrund. Realskolans lärare litade dock inte på folkskollärarnas betygsättning. Därför behövde betygen enligt Wigforss normeras för att fungera som inträdesgrund. Med denna bakgrund infördes normrelaterade prov i slutet av 1930-talet, de så kallade Standardproven. Dessa prov låg till grund för kalibreringen av kraven i läroplanen för den enskilda eleven, och blev därför ett nödvändigt verktyg i utvecklingen av enhetsskolan i Sverige (Lundahl, 2006, 2009). Nationellt utarbetade normrelaterade prov, de så kallade Centralproven, var också en viktig del i avvecklandet av studentexamen som försvann 1968.

I Norge såg utvecklingen devis annorlunda ut vilket bland annat kan förklaras med landets svagare utvecklade vetenskapliga utbildningstraditioner och en starkare skepsis till centralstyre. När politiska reformer i Sverige

institutionaliserades i centrala byråkratiska organisationer, har man i Norge alltid värt sig mot liknande statlig inblandning i lokala instanser. På 1960-talet blev inrättandet av *Forsøksrådet* för *skoleverket* lösningen på den norska skepsisen till centralstyre. Etableringen av *Pedagogisk forskningsinstitutt* (PFI) vid Oslo universitet 1938 hade varit ett viktigt led i satsningen på vetenskapen i uppbyggandet av det norska skolsystemet. Detta institut hade länge en slags monopolställning som vetenskaplig leverantör av skolrelaterad forskning fram till slutet av 1960-talet. I början hade 1930-talets reformpedagogiska idéer från USA och Tyskland stor påverkan på institutets verksamhet, inte minst genom institutets första professor, Helga Eng (Stafseng, 2009).

Eng efterföljdes som institutsledare 1948 av Johannes Sandven, som var mycket inspirerad av det psykometriska ämnesområdet som växte fram i USA och i varierande grad spreds i Europa. Sandvens disputation kan användas som illustration på vilket perspektivskifte som var på gång vid PFI och i norsk pedagogik under denna tid. Som opponent avslutade Eng med att ge uttryck för en förhoppning om att doktoranden i sitt vidare arbete *inte* skulle utveckla sig till en renodlad testpsykolog (Visle, 2009). Det var dock enligt många mening precis det Sandven gjorde. Forsøksrådets nya progressiva pedagogiska riktning representerades genom Sandven vid PFI, beskriven av Telhaug (2003, s. 7) som "Norges Husén". Enligt Tønnessen (2011) etablerades rådet med en förväntning om att experter kunde lösa skolproblemen man stött på. Av samma skäl hade Statens psykologisk pedagogisk institut (SPPI) inrättats i Sverige 1944 och kom fram till 1954 att stå för utvecklingen av Standardprov, då Lärarhögskolan i Stockholm och Torsten Husén tog över den verksamheten (Lundahl, 2009). Standardiserade prov enligt amerikansk testteori kom emellertid inte att etableras i Norge vid denna tid.

Även om amerikanska testidéer fick genomslag i det svenska utbildnings-systemet ska det understrykas att den svenska traditionen skiljer sig markant från den amerikanska genom att det är lärarna själva som bedömer sina elevers prestationer på de nationella proven. Detta grundar sig på att Wigforss envisades med att proven skulle vara ett *stöd* för lärarna i deras bedömning av eleverna; en grundtanke som finns kvar för proven än idag, även om fler syften tillkommit. Av samma skäl var proven i många år inte obligatoriska, utan frivilliga för lärarna att använda sig av (Lundahl, 2009). Om centralt konstruerade prov vinner legitimitet beror alltså dels på hur man lyckas sörja för provens kvalitet, men också hur proven relateras till lärarkårens status och professionssträvanden. Här ser det olika ut i Sverige där fokus legat på att komplettera lärares kompetens och Norge där tonvikten varit att vila på lärares erfarenhet.

PROVENS LEGITIMITET GENOM FÖRANKRING I LÄRARERFARENHETEN

Gemensamt för de svenska nationella proven och de norska examensproven är att de har vuxit fram med förankring i lärarprofessionen. Denna förankring i lärarprofessionen gav provens initiala utveckling nödvändig legitimitet. De första svenska standardproven utvecklades inte i första hand för en central användning och kontroll utan för lokala behov; att hjälpa läraren i sin undervisning. En av upphovsmännen till de svenska standardproven, Frits Wigforss (1886–1953), deklarerade exempelvis tydligt:

Men viktigt är att standardproven få bli vad de är avsedda att vara: en mätsticka i klasslärares hand och ej i en kontrollerande skolmyndighets (Wigforss, 1933, s. 222).

Wigforss var tydligt av uppfattningen att proven skulle utgå från lärarnas perspektiv och utgöra en standardisering av den undervisning de bedrev. Proven skulle inte ersätta lärarnas omdöme utan utveckla det. Standardproven kom dock att centraliseras mer och mer vilket kan förstås som en del av skolans modernisering (se vidare Lundahl, 2009). Detta innebar emellertid att vissa syften med proven fick företräde framför andra. Vissa kvaliteter tillfördes, andra försvann. Bakom de spänningar som finns kring proven står frågan om vad som ska standardiseras och utifrån vems perspektiv dessa standarder bestäms. I potten ligger frågor om lärares professionalisering, elevers lärande, decentralisering och demokrati.

Wigforss klev in i en lång pedagogisk tradition av att bedöma elever och deras prestationer för både summativa och formativa syften där läraren hade haft närmast total autonomi. Wigforss ville vidga denna kontext samt systematisera bedömningarna med inslag av vetenskapligt tänkande. Syftet var dock aldrig att ersätta lärarnas bedömningar med vetenskapliga bedömningar utan att förbättra dem, initialt framför allt för att eleverna skulle kunna få adekvat undervisning.

Det Wigforss gjorde var att han till lärares erfarenhetsbaserade bedömningar tillförde ett så kallat snabbt språk (Lundahl, 2009; Lundahl & Waldow, 2009). Genom att låta ett riksrepresentativt genomsnitt normera lärarnas blick kunde blicken bli mer objektiv. I samma stund skapades också en annan funktion av lärarnas bedömningar. När folkskollärarnas betyg standardiserats med hjälp av standardproven blev det möjligt att avskaffa inträdesexaminationen, det vill säga de prov och förhör realskolans lärare genomförde med de folkskoleelever som ville läsa vidare. Förvaltningen av skolan blev i sig själv därmed mer effektiv; snabbare (Lundahl, 2008). Inte minst fick

folkskollärarnas betyg en legitimitet hos mottagande realskollärare och hos allmänheten (Landahl & Lundahl, 2013). Vad Wigforss således åstadkom var en brygga mellan två användningar av prov. Lärare kunde använda dem av pedagogiska skäl, och samhället kunde använda dem för att effektivisera sitt utbildningsväsen och legitimera dess modernisering.

Motiven för de svenska standardproven kom successivt att övergå från att ha sin grund i lärarerfarenheten och den pedagogiska praktiken till att utgå från den vetenskapliga praktiken, när arbetet med dem hamnar under Skolöverstyrelsens ansvar och den faktiska konstruktionen av proven förläggs till forskningsinstitutioner under 1950-talet (Lundahl, 2009). Hela förfarandet med antagning till nästa utbildningsnivå kunde nu konceptualiseras som en objektiv materia. Folkskollärarna gillade naturligtvis inte riktigt detta (Lundahl, 2006). Deras bedömningar av vad som var tillräckliga kunskaper hade utmanats. Men som Porter (1995) skriver om liknande utveckling i USA – lärarna var mestadels ogifta kvinnor och hade inte riktigt det professionella självförtroendet att sätta sig upp mot ”standardiseringens män”. Men även på gymnasiet, som inte på samma sätt som folkskolan fått sin betygsättning ifrågasatt, kom standardiserade prov att utmana lärarnas bedömningar vid slutet av 1960-talet (Lundahl, 2009). Genom att hävda att lärarna skulle få mer tid till undervisningen om staten skötte bedömningen kunde proven dock erövra legitimitet, vilket vi snart återkommer till.

Arbetet med att introducera de svenska statliga proven pågick med ökad intensitet under flera decennier efter andra världskriget. Den norska lärarprofessionen hade under samma period ”full kontroll” över de nationella kunskapsbedömningarna i och med deras roll som censorer vid examen artium. Detta är en del av förklaringen till hur statens inblandning i skolans bedömningar genom de svenska nationella proven kunde öka allt mer också under kommande decennier, trots att de blev allt mer abstrakta i förhållande till vad lärare gör dagligdags, medan de norska standardproven som på försök hade implementerats 1968 möttes av ett mycket starkt motstånd både bland lärare och i opinionen i Norge och aldrig blev riktig institutionaliserade.

PROVENS LEGITIMITET SOM OBJEKTIVA MÄTINSTRUMENT

I sin analys av psykologins språk skriver Kurt Danziger: ”[s]trictly speaking, there are no ‘raw’ data in science. Whatever forms the subject of scientifically informed perception is already classified in some way” (Danziger, 1997). Liksom i till exempel England, Tyskland och Frankrike började Sverige och Norge att använda standardiserade mätningar för att organisera elever med särskilda behov i anstalter eller i särskilda klasser i början av 1900-talet. Specialklasser och institutioner för psykiskt handikappade hade förvisso

funnits länge men med psykologiska test blev placeringen av barn knuten till psykologiska kategorier och standardiserade metoder för utredning. I Sverige introducerades intelligenstest av Gustaf Jeaderholm i en översättning och utprövning 1914 (se vidare Axelsson, 2007). På slutet av 1920-talet uppstod även i Norge ett starkt intresse för intelligenstestning bland psykiatriker och progressiva pedagoger. Bland de mest tongivande var Anna Sethne och Bernhof Ribsskog, som var en del av det som kallades "Oslo-sirkelen".

Många av idéerna bakom de standardiserade grupptesten kom från Amerika. Svenska testutvecklare blev tränade under andra världskriget för att konstruera och genomföra grupptester. Efter kriget blev många av de psykometriker som hade arbetat på militärpsykologiska institutet utan arbetsuppgifter. Flera psykometriker utomlands och i Sverige kom i stället att hamna inom utbildningssektorn. I Sverige blev, som vi antydde ovan, den stora enhetsskolereformen en perfekt träningsplats och "laboratorium" för utveckling av standardiserade tester och beräkning av korrelationer och avvikelser (se Härnquist & Husén, 2000). Utbildningssektorn erbjöd den pedagogiska vetenskapens mätpsykologiska gren en fristad där den kunde utvecklas mot större teoretisk konsekvens och metodologisk färdighet, och därmed erhålla en allt högre legitimitet (se Danziger, 1990; Porter, 1995).

Som vi nämnt kom de svenska standardprovns andra utvecklingsfas att ledas av Torsten Husén vid Lärarhögskolan i Stockholm. Husén var utbildad just vid militärpsykologiska institutet och det arbete han ledde uttrycker med tydlighet hur prov, provkonstruktion och resultatanvändning underordnats vetenskapliga strukturer. Husén och hans kollegor framhöll förvisso att proven var viktiga i den pedagogiska praktiken, men de ville eliminera lärarens inflytande över resultaten. Därtill ville de mäta endast det säkert mätbara. Ett viktigt argument för dem var att deras standardprov i sig själva var effektiva. De skulle gå snabbt att ta fram dem, genomföra dem och inte minst att rätta dem. Målet var att tillhandahålla ett instrument för att differentiera elever som i någon mening påverkade skolans arbete så lite som möjligt (Husén, Björnsson, Edfeldt & Henrysson, 1956). Ett sätt att legitimera centralt konstruerade prov var att framhålla dess effektivitet jämfört med lärares egna provtraditioner. Vid införandet av centralproven i gymnasieskolan var ett huvudargument att tiden i skrivsalarna skulle mer än halveras (SOU 1963:42; Lundahl, 2009). Detta välkomnades av både lärare och elever som var tyngda under "provbördan" men det innebar också att kommunikationen om och med bedömningar bland lärare minskade (Lundahl, 2006), och därmed försvann också flera självreglerande och/eller fortbildande inslag i bedömningsfrågor.

I Norge hade det parallellt pågått ett liknande arbete med att få till stånd en säkrare betygssättning. En av de mest inflytelserika aktörerna inom utbildningssektorn, bland annat som rektor vid Oslo Katedralskole från 1927 till 1937, var Hans Eitrem. Hans analys av betygssättningen vid examen artium under perioden 1920-1924 var startskottet för ett ökat fokus på jämförbarhet i lärares betygssättning (Lysne, 1999). Norge hade precis som Sverige deltagit i det banbrytande internationella samarbetet om examens- och betygsfrågor, *The international examination inquiry*, framför allt genom Kay Piene och Olav Sundet, som 1939 publicerade en undersökning om *Standpunkt-karakterer og eksamens-karakterer ved skriftlig middelskoleeksamen*. Ribsskog arbetade också med dessa frågor och analyserade betygssättningen i några *folkeskoler* i Osloområdet. I en studie från 1938, lät han bland annat olika lärare värdera samma elevers arbeten och upptäckte ganska stora avvikelser mellan lärarnas betygssättning, med en tendens mot att de höga betygen överanvändes (Lysne, 1999).

Vid PFI kom under ledning av Sandven psykometriska tester efter hand att dominera institutets verksamhet, men Sandven hade inte den koppling till lärarprofessionen som Wigforss hade i Sverige och detta försvagade provens och sedan även institutets legitimitet. Det uppstod efterhand en stor frustration och motstånd mot Sandvens ståndaktighet och bristande pedagogiska perspektiv (Visle, 2009). Sandvens kompromisslösa projekt med att upprätta en vetenskaplig psykometrisk disciplin, kulminerade i kraftfulla konfrontationer när positivismstriden och studentupproren härjade i samhällsvetenskapen i slutet av 1960-talet (Helsvig, 2005).

På 1970-talet växte det som Telhaug benämner som nyradikalismen fram som en inflytelserik rörelse. Nyradikalismen hade en liberal och marxistisk grund, men upprepade mycket av kritiken som på en kristen värdegrund framfördes mot den etablerade pedagogiken på 1940 och 1950-talet. Erling Kristvik raljerade enligt Telhaug (1984) över "kvotient- og koeffisient-utrekningar" vid PFI, och beklagade sig över att ämnet dominerades av kvantitativa metoder och psykometriska studier (s. 167). Hans Skjervheim (1972) varnade för att göra den naturvetenskapliga modellen absolut, eftersom det kunde leda till ett "instrumentalistisk mistak". Detta väckte stor begeistring bland akademiker och andra intellektuella eftersom det gav näring till den redan så omfattande kritiken av PFI.

Kort tid efter att Sandven ofrivilligt tvingades avgå som institutsledare 1972 klyvdes ämnet i två delar. Från 1974 kunde studenterna välja ett oppositionellt *sosialpedagogisk studiealternativ*. En av de centrala aktörerna bakom det socialpedagogiska alternativet, Eva Nordland, slog fast att det inte var möjligt att beskriva en människa med hjälp av mätningar, experiment eller tester (Telhaug, 1984).

Maktkampen på PFI ger uttryck för hur 1970-talets politiska stämning fick mycket större genomslag i norsk bedömningspolitik än vad som var fallet i Sverige. Även om man i efterkrigstiden närmast kopierade svenskarna i uppbyggandet av enhetsskolan (Sejersted, 2005), så höll man fast vid den europeiska examenstraditionen och lät sig inte bli lika "amerikaniserade" som svenskarna vad det gäller bedömningsformer. Den pedagogiska psykologin i Norge lät sig inte besegras helt, men som skolhistorikern Jarning (2009) påpekar är det intressant att notera hur svensk empirisk psykologisk baserad skolforskning kom att utveckla mer samhällsvetenskapliga analyser av undervisningsprocesser, medan norsk forskning höll sig mer fast vid psykologisk individualism.

I Sverige fanns det förstås också motstånd mot social ingenjörskonst, instrumentalism och vetenskaplig positivism, men psykometrisk forskning och psykometriska prov hade tydligt en starkare legitimitet än i Norge. Det hänger sannolikt ihop med att staten sett till att proven i Wigforss anda skulle uppfattas som hjälpmedel för lärarna i undervisning. Läger vi till det faktum att proven gav de relativa betygen ett skimmer av objektivitet är acceptansen för proven i Sverige än lättare att förstå. Med Danziger (1997) kan man säga att skillnaden mellan utvecklingen av objektiva test i Sverige och i Norge visar att en viktig förutsättning för psykometrin att lyckas få genomslag i skolväsendet var förmågan att inte bara knyta an till språkliga kategorier utan även till professionella kategorier.

NATIONELLA BEDÖMNINGSINSTRUMENTS ROLL

I STANDARDISERINGEN (NORMERINGEN) AV LÄRARNAS BETYGSÄTTNING

En bärande idé med nationella bedömningsinstrument är att de ska fungera som norm för ett lands lärare. I Norge infördes principer om normalfördelning av betyg i och med 1939 års Normalplan. Så som det senare blev i Sverige utgick man ifrån Gausskurvan, som blivit etablerad praxis i den internationella testmiljön. I den norska läroplanen 1939 bestämdes normalfördelningen på följande sätt: Særs godt, 4%; Meget godt, 24%; Godt, 44%; Nokså godt, 24%; Lite godt, 4%. Det preciserades att den vägledande normalfördelningen skulle ha giltighet bara för ett större antal betyg, exempelvis flera tusen betyg, och att det i de flesta klasser på cirka 30 elever alltid skulle innehålla avvikelser åt antingen det ena eller det andra hållet (Lysne, 1999). Samma princip introducerades i Sverige under 1940-talet men med den intressanta skillnaden att "de flera tusen" betygen konstruerades av

standardprov medan de i Norge var imaginära. Tanken både i Norge och i Sverige var alltså att erbjuda en norm för läraren att positionera *klassen* efter, även om många lärare verkade använda riksnormen också på individnivå (jämför Andersson, 1999).

Normerings- och bedömningsprinciperna bakom normalfördelningen i Norges Normalplan 1939 är svår att beskriva med dagens bedömningsterminologi. I beskrivningen av kraven för det bästa betyget (Sers godt) lades det stor vikt på hur sällan sådana betyg skulle utdelas. Medan de andra betygsbeskrivningarna (Mykje godt, Godt, Nokså godt, Lite godt) var kortfattade (mellan 7 och 17 ord) och begränsade sig till distinktioner i förhållande till medelnivå (med ett tillägg om att Skolstyrelsen skulle avgöra avvikelser från normalfördelningen på det lägsta betyget), ges det en utförlig beskrivning (212 ord) av villkoren som måste uppfyllas för att få toppbetyg (Lysne, 1999).

En viktig funktion med normalfördelningsnormen var att begränsa inflationen i betygssättningen. Nationella prov får därmed legitimitet genom att de kan bidra till ökad rättvisa. Det speciella – särskilt sett med svenska och amerikanska ögon – är hur man kunde införa en normalfördelningsprincip i Norge utan att det förankrades i standardiserade prov. I vilken grad detta reflekterade resursproblem, bristande kompetens och förståelse för testprinciper som ansågs vara elementära på andra platser, eller helt andra grunder, kan vi inte dra några slutsatser kring.

Betygen i Norge skulle följa normalfördelningen men baserat på att lärare i sig själva var skickliga nog att avgöra ungefär var gränserna gick. I Sverige tog utvecklingen en annan riktning och vetenskapliga tester skulle vara det som säkerställde kvalitet, framför allt i bedömningen – genom att minimera betydelsen av lärarens bedömarkompetens. Lärarna skulle göra det de var bra på, att undervisa, medan vetenskapsmännen tog ”tempen” på eleverna, ställde diagnos och meddelade lärarna utfallet, argumenterade Husén med kollegor i sin bok Standardproven (1956). Inslag av autenticitet och subjektivitet skalades bort, och kopplingen till undervisningens innehåll var svag. Husén och hans forskningsgrupp skrev:

Flervalstypen är överlägsen i flera viktiga avseenden. Den eliminerar lärarens uttal, dialekt, hastighet i talet, kontakt med klassen och andra okontrollerbara faktorer som inverkar vid diktamenstypen. Man slipper sätta fel för oläsligt skrivna ord, ändringar, otydligheter, utbytta ord och böjningsformer m.m. Dessutom går både prövningen och rättningen betydligt fortare än vid diktamenstypen (Husén, Björnsson, Edfeldt & Henrysson, 1956, s. 45).

Med de första svenska standardproven togs en hel del ansvar för bedömningen bort från den svenska lärarkåren. Detta förminskande av lärarbedömningarnas betydelse behövde förstås legitimeras. Som vi har beskrivit ovan tjänade i Sverige den historiska kopplingen mellan dessa prov, som bryggan mellan en pedagogisk användning och en administrativ användning en viktig roll. Känslan av att betygen blev mer objektiva och rättvisa genom externa test hade betydelse (Lundahl, 2006).

I båda länder använde man relativ betygsättning, men i Norge utan att man använde standardiserade prov vilket man i Sverige sett som de relativa betygens förutsättning. Detta indikerar att det i båda länderna pågick aktivt legitimerings- och "dislegitimerings-arbete". Om lärares betyg inte anses behöva standardiseras behövs ju inte heller prov som åstadkommer detta.

En förklaring till att den normalfördelade betygssättningen i Norge inte behövde legitimeras med vetenskapliga prov är att Norge som aldrig haft en parallellskola heller inte hade de professionsstrider som fanns mellan folkskollärare och realskollärare i Sverige (se till exempel Lundgren, 2010), där folkskollärarna behövde legitimera betygen gentemot realskolan. En annan faktor är att Norge var (och är) en ung nation som inte hade tillräcklig kompetens att utveckla "objektiva prov", varför lärares bedömningar så att säga aldrig konkurrensutsatts. Detta problem såg man senast också vid införandet av "nasjonale prøver" 2004, när myndigheterna plötsligt blev tvungna att ta ett års "provepauz" medan de utredde bättre provsystem (Lie, 2007; Tveit, 2007). Poängen är således att nationella prov inte helt utan vidare kan ersätta lärares prov, utan de måste legitimeras, så som mer reliabla eller mer valida. I Sverige låg fokus på reliabilitet, i Norge har bristen på validitet bidragit till att psykometriska prov haft svårt att få legitimitet.

PROVENS ROLL I MÅL- OCH RESULTATSTYRNING

Centralt konstruerade prov har lättare att få legitimitet om validiteten är hög. Provns validitet är dock nära förbunden med hur de används och med vilka effekter (Messick, 1989). Centralt i debatten om proven i Sverige har varit dess användbarhet mot olika syften och om provns reella konsekvenser. De psykometriska proven uppfattades exempelvis under 1970-talets slut och under 1980-talet ha gått väl långt i sin strävan efter vetenskaplig objektivitet (Lindblad, 1991). Lärare och elever började ifrågasätta meningsfullheten med att genomföra den typen av prov.

Samtidigt pågick ett utvecklingsarbete med att integrera proven i nationella utvärderingar, det vill säga använda dem som skolans resultat och inte längre bara elevens. Detta innebär att proven får ett nytt syfte utöver att främja individuell rättvisa. Debatterna om skolan under 1970- och 80-talen

i Sverige väckte ett behov av att synliggöra och följa upp resultat. Det fanns starka indikationer på att skolan inte levde upp till sina anspråk och Riksrevisionsverket hade kraftfullt kritiserat Skolöverstyrelsen för att inte leverera systematiska underlag och analyser (RRV, 1980). Staten behövde resultatbilder men var också medveten om att utvärderingar och prov styrde på ett sätt som var svårt att förena med nya ambitioner i läroplanen om decentralisering av ansvar och elevcentrerad undervisning.

Lösningen blev en typ av *marknadsmodell* där staten angav övergripande mål för kommunerna att arbeta mot. Kommunerna skulle sedan själva utvärdera om målen nåtts och förse staten med resultat. Parallellt skulle staten ta fram stödmaterial och i befogade fall göra egna utvärderingar. I den mån resultat rapporterades in visade det sig dock att de inte producerats under likvärdiga former. Skolverkets granskning av kommunernas kvalitet intensifierades allt mer under slutet av 1990-talet och början av 2000-talet, till en början genom att man krävde in detaljerade kvalitetsredovisningar där betygsstatistik och provresultat var några av resultatmått (Nytell, 2006).

År 2003 införs på Skolverket en avdelning för kvalitetsinspektion, vilken sedan blir en helt egen myndighet 2008. Skolinspektionen har under sina hittillsvarande verksamhetsår skärpt granskningen av både betygssättning och bedömning av nationella prov, då dessa resultatmått spelar en så stor roll för att systemet ska fungera (Gustafsson & Erickson, 2013). Trots ökade behov av "säkra" resultat har dock knappast någon "fortbildning" av lärarkåren i bedömningsfrågor skett fram till 2010-talet (jämför Wermke, 2013). Då, i samband med Lgr 2011 och 2012 års betygssystem intensifierar Skolverket utvecklingen av stödmaterial kring bedömning. Om lärares betyg eller deras bedömning av nationella prov, där lärarna i Sverige själva står för rättningen, inte anses likvärdiga eller reliabla, befinner sig hela utbildningssystemets legitimitet i farozonen, då det bygger på möjligheten att göra informerade val i kombination med kompensatorisk resurstilldelning (jämför Vlachos, 2012).

Både i Sverige och i Norge kom utredningar under början av 2000-talet som bedömde att tolkningsutrymmet var för stort för lärarna. I Norge är det bästa mått man har på lärarnas betygssättning en jämförelse mellan skolornas betygssättning och examensbetygen. Sådana jämförelser har påvisat stora skillnader (Kristensen 1999; Kommunerevisjonen 2013; Prøitz & Spord Borgen, 2010). I Sverige har Skolinspektionens omrättning av nationella prov skapat mycket debatt i bland allmänheten. Gustafsson och Erickson (2013) menar att Skolinspektionen använder en "skolinspektionslogik" och inte en forskningslogik i sin omrättning. Gustafsson och Erickson menar att skolinspektionen gjort ett tankefel när de utgått ifrån att externa bedömare oavsett kompetens kan göra objektiva bedömningar av proven, och att dessa bedömningar underförstått skulle vara säkrare än den ursprungliga

lärarens. Att detta inte blev tydliggjort när Skolinspektionen fällde sin ”dom” över lärarnas och skolornas bedömningar, kan hota provens legitimitet som bedömningsinstrument och lärarnas legitimitet som bedömare.

En ökad precisering av mål och krav i styrdokumentet bedöms kunna ge en ökad likvärdighet i bedömningen av elevers kunskaper och graden av målpåfyllelse. Den bedöms också underlätta uppföljningar på verksamhetsnivå. Den svenska regeringen införde därför tydligare och mer lättillgängliga läroplaner med nya kursplaner. För att ytterligare stärka uppföljningen av elevernas kunskaper införde regeringen 2011 nationella kunskapskrav i årskurs 3, 6 och 9 (Prop. 2008/09:87) med betyg från årskurs 6 och uppåt. En av de stora utmaningar *policy makers* står inför är hur de ska kunna legitimera ökad styrning av lärares undervisning och av deras bedömning. De huvudalternativ som står till buds är ökade krav och införande av system för *accountability* eller utbildningssatsningar. Medan det förra kan upplevas som alltför tvingande och har låg legitimitet hos lärare, kan det senare upplevas som alltför frivilligt och har låg legitimitet hos politiker och allmänhet. Följande ledare i Dagens Nyheter kan illustrera detta dilemma:

Betygssystemets så kallade kunskapskrav utgör inte det ankare som skolbyråkrater, politiker och ibland även lärare låtsas att de är. Rättvisa betyg uppstår inte i mötet mellan elevers prestationer och statliga betygs-kriterier. Betyg sätts av människor med egna kunskaper, erfarenheter och värderingar (Dagens Nyheter, 2013-10-30).

Det råder närmast ett ”dödläge” mellan förvaltningsbehov och professionsbehov. I ett sådant dödläge har den bedömningsinriktning som kallas formativ bedömning eller bedömning för lärande varit en kompromiss, vilket vi tror förklarar en del av dess genomslag.

DISKUSSION: PROVENS FÖRLORADE LEGITIMITET KOMPENSERAS MED NYA FORMATIVA SYFTEN

När bedömningsdata används för pedagogiska förändringar tycks det alltid uppstå vissa spänningar kring 1) hur data ska samlas in, 2) hur data ska standardiseras och, 3) hur data kan användas (Lundahl & Lundahl, 2013). Att använda data i utbildningsreformer är ett konstant överläggande mellan olika aktörer om validitet och reliabilitet – det handlar om att kalibrera data med våra epistemologiska och ontologiska föreställningar såväl i vetenskap som i vardag. Danziger skriver: “in the end, behind every score there is a judgment, a judgment that is a measure of, for example: intelligence, talent, school result” (Danziger, 1997, s. 83).

Genom historien kan vi se två krafter som verkar mot varandra i kalibreringen av anspråk och användning för bedömningar. Å ena sidan har vi skolan som institution där inte minst lärarna behöver göra tillitsfulla och pedagogiskt användbara bedömningar för att legitimera sin roll som utbildare, å andra sidan har vi staten och politikerna som inför medborgarna behöver visa att skattefinansierade verksamheter håller hög kvalitet. Här finns vissa intressekonflikter. I skolan under 1800-talet och 1900-talets första hälft hade lärare stora rättigheter att avgöra vilka de ville ta emot och skicka vidare till fortsatt utbildning genom examensförfaranden. Medborgarna hade däremot inte särskilt självklara rättigheter till utbildning. Eleven var tvungen att visa sitt värde i skolinstitutionen. Genom att politikerna tog på sig att ge allt fler individer möjligheter att studera allt längre, började de också styra över rätten till vidare utbildning.

Vi kan från andra halvan av 1900-talet till idag se hur lärares rätt att själva definiera vad som ska bedömas och hur, kringskurits till fördel för medborgarnas 1) rätt till utbildning, 2) rätt till stöd under utbildning, 3) valfrihet inom utbildning. Medan skolan som institution ursprungligen använde bedömningar för att säkra sin egen överlevnad har det blivit den enskilde individens möjligheter att använda skolan för att förverkliga sina behov och önskemål som nu står i centrum.

I ljuset av detta resonemang är det intressant att reflektera över skillnaderna mellan Sverige och Norge. I Norge har de nationella examinationerna upprätthållit legitimiteten genom att det är professionen själv som kontrollerar proven. Att få vara censor räknas som ett erkännande av lärarnas kompetens, och ledningen vid skolor önskar att deras lärare ska vara censorer för att få nytta av den kompetenshöjning som följer med att träna sig till censor och delta vid censorsbedömningarna. Proven har begränsad roll i att kvalitets-säkra utbildningen. När man i Norge införde "nasjonale prøver" med det primära syftet att kontrollera kvaliteten på skolornas (och därigenom lärarnas) utbildning, fick man stora problem som följd av bristande kompetens i utformandet av dessa prov. Men det starka motstånd proven möttes av från lärarhåll kan *också* ses i ljuset av dess syfte att kontrollera. Att dessa två syften i Sverige ingår i ett och samma provsystem utgör ett fundamentalt hot mot provens legitimitet i lärarprofessionen. De norska examinationsproven lyckas i större grad att undvika sådan kritik eftersom de inte i samma utsträckning har en central roll i utbildningssystemets mål- och resultatstyrning.

Inledningsvis frågade vi vad som händer om samhället inte längre litar på lärares bedömningar? Skolinspektionen och andra institutioner som ingår i det moderna förvaltningssystemet bidrar till att sätta ökad press på provens legitimitet i opinionen. Men som Bourdieu och Passeron visade i Frankrike kan provens legitimitet upprätthållas även utan legitimitet i opinionen,

såvitt det är en funktionell del av utbildningssystemets autonomi och har legitimitet i professionen. Lärare har historiskt sett haft en känsla av att deras bedömningar bidrar till sitt ämnes kunskapsreproduktion, reproduktionen av duktiga lärare eller att bedömningen fyller pedagogiska funktioner. Det är lätt att förstå att en lärare kan känna stolthet över den utveckling hon bidragit till hos ett specifikt barn. Men det är sannolikt svårare att se sitt lilla bidrag till ett stort dataunderlag i en kvalitetsredovisning. När det blir ett alltför stort glapp mellan lärares vardagsbedömningar och de bedömningar som hålls fram på nationell nivå uppstår en legitimitetssvikt. Därför är det begripligt att både de svenska nationella proven och de norska "nasjonale prøver" nu legitimeras av utbildningspolitiker och myndigheter genom hänvisningar till internationell bedömningsforskning och att man då återoppar bedömningens potential för lärandet. Det finns förstås en risk att även denna legitimeringsstrategi undergräver lärares professionalisering i bedömningsfrågor.

Denna artikel har visat att utbildningspolitiker, skolmyndigheter och bedömningsforskare har svårt att upprätthålla legitimiteten i ett provsystem som utformades i en helt annan tid där uppmärksamheten om provens funktioner och kvaliteter var långt mindre än de är i dagens målstyrda utbildningssystem. Vi tror att en balans mellan nationella bedömningsinstruments legitimitet och lärares professionalisering borde bygga på att lärare genom utbildning återerövrar autonomi över bedömningarna genom att sätta elevers och lärares lärande i fokus. Dock på ett sätt som kan tillfredsställa andra professioners och samhällets behov av uppmätta skolresultat. Från statsmakternas sida handlar det om att stödja lärarna i denna kompetensutveckling och om att inte utnyttja lärares bedömningar för sådant de inte är lämpade till. Men också att staten utvecklar uppföljningsinstrument som förmår ta vara på mer kvalitativa resultatbilder och som i mindre grad än för närvarande intervenerar med skolans dagliga verksamhet. Den historia vi tecknar visar hur fler och fler syften tillkommer de nationella proven i Sverige medan de nationella examina i Norge i huvudsak handlat om en sak, att säkerställa certifieringen av elevernas kunskaper i slutet av skolan. Det faktum att svenska nationella prov härigenom blivit obligatoriska och att rättningen blivit allt mer externt kontrollerad torde haft stor negativ inverkan såväl på provens legitimitet bland lärare och på lärarnas status som bedömare. De kvaliteter som finns i dagens svenska nationella prov skulle sannolikt komma bättre till sin rätt om lärarna fick ökat självbestämmande över dem, tid att arbeta med bedömningsstöd och sambedömning, samt att ett mindre antal, men väl valda, frågor användes för urvalsutvärdering.

REFERENSER

- Andersen, J.A. (2009). *Organisasjonsteori. Fra argument og motargument til kunnskap*. Oslo: Universitetsforlaget.
- Axelsson, T. (2007) *Rätt elev i rätt klass: Skola, begåvning och styrning, 1910-1950*. Doktorsavhandling. Linköping: Linköping studies in arts and science, 379.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N., (2010). Validity in teachers' summative assessments, *Assessment in Education: Principles, Policy & Practice*, 17(2), 215–232.
- Bourdieu, P. och Passeron, J.-C. (1970/1990). *Reproduction in Education, Society and Culture*. London/Thousand Oaks/New Delhi: Sage Publications.
- Carlgren, I. (1986). Lokalt utvecklingsarbete. *Acta Universitatis Gothoburgensis. Göteborgs universitet: Studies in Educational Sciences*, 56.
- Carlgren, I. (1999). Professionalism and teachers as designers. *Journal of curriculum studies*, 31(1), 43–56.
- Carson, J. (2006). *The measure of merit. Talents Intelligence, and Inequality in the French and American Republics, 1750-1940*. Princeton, NJ: Princeton University Press.
- Collins, S., Reiss, M. & Stobart, G. (2010). What happens when high-stakes testing stops? Teachers' perception of the impact of compulsory national testing in science of 11-year-olds in England and its abolition in Wales. *Assessment in Education: Principles, Policy & Practice*, 17(3), 273–286.
- Dale, E.L. (2008). *Fellesskolen – reproduksjon av sosial ulikhet*. Oslo: Cappelen akademisk forlag.
- Dale, E.L. (2009). Kritisk pedagogikk ved PFI. *Norsk Pedagogisk Tidsskrift*, 93(6).
- Dagens Nyheter (2013-10-30). <http://www.dn.se/ledare/huvudledare/laggned-skuggspelet/?brs=m>
- Danziger, K. (1990). *Constructing the subject: historical origins of psychological research*. Cambridge: Cambridge University Press.
- Danziger, K. (1997). *Naming the mind: how psychology found its language*. London: SAGE.

- Forsøksrådet for skoleverket (1969). *Standardiserte prøver i skolen. Forsøk og reform i skolen – nr 16*. Oslo: Universitetsforlaget
- Florin, C. (2010). Lärarnas professionalisering. Nedladdad 22.05.2014 från: http://www.lararnashistoria.se/sites/www.lararnashistoria.se/files/artiklar/Lärarnas%20professionalisering_0.pdf
- Gustafsson, J.E. & Erickson, G. (2013). To trust or not to trust? Teacher marking versus external marking of national tests. *Educational Assessment, Evaluation and Accountability*, 25(1), 69–87.
- Helsvig, K. (2005). *Pedagogikkens grenser. Kampen om norsk pedagogikk ved Pedagogisk forskningsinstitutt 1938-1980*. Oslo: Abstract forlag.
- Husén, T., Björnsson, C. H., Edfeldt, Å. W & Henrysson, S. (1956). *Standardproven. En redogörelse för konstruktion och standardisering*. Stockholm: Almqvist & Wiksell/ Geber.
- Husén, T. & Härnquist, K. (2000). *Begåvningsreserven. En återblick på ett halvsekel forskning och debatt*. Stockholm: Föreningen för svensk undervisningshistoria, vol. 193.
- Jarning, H. (2009). En uregjerlig disiplin i søtti år – PFI og pedagogikk mellom nabofag, profesjoner og bevegelser. *Norsk Pedagogisk Tidsskrift*, 93(6), 507–513.
- Johannisson, K. (2006): Hur skapas en diagnos? Ett historiskt perspektiv. I G. Hallerstedt (red.), *Diagnosens makt. Om kunskap, pengar och lidande* (ss. 29–41). Göteborg: Daidalos.
- Kommunerevisjonen (2013). *Standpunktkarakterer i videregående skole – likebehandles elevene?* Oslo: Oslo kommune kommunerevisjonen.
- Lawn, M. (Ed.) (2008): *An Atlantic Crossing? The work of the International Examination Inquiry, its researchers, methods and influence*. Oxford: Symposium Books.
- Lie, S. (2007). Evalueringen av de nasjonale prøvene og hva vi kan lære av dem. I S. Tveit (red.), *Elevvurdering i skolen - grunnlag for kulturendring* (ss. 85–92). Oslo: Universitetsforlaget.
- Lindberg, V. (2005). Svensk forskning om bedömning och betyg 1990–2005. *Studies in Educational Policy and Educational Philosophy*. E-tidsskrift 2005:1. Internet: www.upi.artisan.se nedladdad 2006-08-18.
- Lindberg, V. & Forsberg, E. (2010). *Svensk forskning om bedömning. En kartläggning*. Vetenskapsrådets rapportserie 2:2010. Stockholm: Vetenskapsrådet.

- Lindblad, T. (red.) (1991). *"Allt är relativt". Centralt utarbetade prov i svenska, matematik, engelska, franska, tyska, fysik och kemi – beskrivning, analys, förslag*. Rapport nr 1991:03, Institutionen för pedagogik, Göteborgs universitet.
- Lundahl, C. (2008). Inter/national assessments as national curriculum: the case of Sweden. In M. Lawn (Ed.), *An Atlantic Crossing? The work of the International Examination Inquiry, its researchers, methods and influence* (pp. 157–180). Oxford: Symposium Books.
- Lundahl, C. (2009). *Varför nationella prov? – Framväxt, dilemma, möjligheter*. Lund: Studentlitteratur.
- Lundahl, C. (2010). Varför bedömer vi kunskap? I C. Lidberg, U. P. Lundgren & R. Säljö (red.), *Lärande, skola, bildning. Grundbok för lärare* (ss. 255–292). Stockholm: Natur och kultur.
- Lundahl, C. (2011). *Bedömning för lärande*. Stockholm: Nordstedts förlag.
- Lundahl, C. & Petterson, D. (2010). Den svenska skolens resultat. Från standardprov til PISA. I E. Elstad och K. Sivesind. *PISA - sannheten om skolen?* (ss. 222–243). Oslo: Universitetsforlaget.
- Lundahl, C. & Waldow, F. (2009). Standardisation and 'quick languages': the shape-shifting of standardized measurement of pupil achievement in Sweden and Germany. *Comparative Education*, 45(3), 365–385.
- Ludvigsen, K., Lundahl, C. & Ydésen, C. (2013). Creating an Educational Testing Profession in Norway, Sweden and Denmark, 1910-1960. *European Educational Research Journal*, 12(1), 120–138.
- Lysne, A. (1999). *Karakterer og kompetanse. Stridstema i norsk skolehistorie*. Haslum: AVA forlag.
- Lysne, A. (2004). *Karakterer og kompetanse. Kampen om skolen*. Haslum: AVA forlag.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement*. 3. Utgåvan, 1993, (pp. 13–103). Phoenix, AZ: The Oryx Press.
- Mickwitz, L (2011). *Rätt betyg för vem? Betygsättning som institutionaliserad praktik*. Licentiatavhandling Stockholm: Stockholms universitet. Tillgänglig på Internet: <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-54614>
- Nytell, H. (2006). *Från kvalitetsidé till kvalitetsregim. En studie om statlig styrning av skolan*. Doktorsavhandling. Uppsala universitet: Acta Universitatis Upsaliensis.

- Persson, S. (2008). *Läraryrkets uppkomst och förändring. En sociologisk studie av lärares villkor, organisering och yrkesprojekt inom den grundläggande utbildningen i Sverige 1800-1900*. Doktorsavhandling. Göteborgs universitet: Göteborg
- Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and in public life*. Princeton, NJ: Princeton University Press.
- Prop. 2008/09: 87. Tydligare mål och kunskapskrav - nya läroplaner för skolan. Regeringens proposition: <http://www.regeringen.se/content/1/c6/11/72/69/514f5f8a.pdf>
- Prøitz, T. & Spord Borgen, J. (2010). *Rettferdig standpunktvrdering – det umuliges kunst (?)*. Oslo: NIFU STEP.
- RRV (1980): *Skolöverstyrelsens utvärdering: provverksamhet: revisionsrapport*. Stockholm: RRV.
- Scriewer, J. (Ed.) (2012). *Discourse formation in comparative education*. 4. utgåvan. Frankfurt am Main: Peter Lang Verlag.
- Sejersted, F. (2005). Sosialdemokratiets tidsalder. Norge og Sverige i det 20 århundre. *Norge og Sverige gjennom 200 år*. Oslo: Pax forlag.
- Slagstad, R. (1998). *De nasjonale strateger*. Oslo: Pax forlag.
- Stafseng, O. (2009). Helga Eng og hennes pedagogikkfag 1938 – mellom tradisjon og arvefølge. *Norsk Pedagogisk Tidsskrift*, 93(6), 433–442.
- Skjervheim, H. (1972). Det instrumentalistiske mistaket. I J. Hellesnes & G. Skirbekk (red.), *Mennesket*. Oslo: Universitetsforlaget
- Skolverket (2014): Betygshistorik. Nedladdad 22.05.2014 från: <http://www.skolverket.se/bedomning/betyg/betygshistorik-1.46885>
- SOU 1963:42. 1960 års gymnasieutredning (1963). *1960 års gymnasieutredning. Ett nytt gymnasium*. Stockholm: Ecklesiastikdepartementet.
- Stortingsmelding nr 16 (2006-2007). *Og ingen stod igjen. Tidlig innsats for livslang læring*. Oslo: Kunnskapsdepartementet.
- Telhaug, A.O. & Mediås, O.A. (2005). *Grunnskolen som nasjonsbygger: fra statspietisme til nyliberalisme*. Oslo: Abstrakt forlag.
- Tveit, S. (2007). Elevvurdering i Kunnskapsløftet. I H. Hølleland (red.), *På vei mot Kunnskapsløftet* (ss. 281–302). Oslo: Cappelen akademiske forlag.

- Tveit, S. (2014a). Educational Assessment in Norway. *Assessment in Education: Principles, Policy & Practice*, 21(2), 221–237.
- Tveit, S. (2014b). Constructing Legitimate Meritocracies: Comparing policies for employing external instruments ensuring comparability of high-stakes teacher judgments in Scandinavian secondary education. Paper presenterad vid *The European Conference of Educational Research*, Porto, 2. September 2014.
- Tønnesen, L.K.B. (2011). *Norsk utdanningshistorie. En innføring med fokus på grunnskolens utvikling*. Oslo: Fagbokforlaget.
- Visle, L. (2009). Den Sandvenske “gullalder” på 50- og 60-tallet. *Norsk Pedagogisk Tidsskrift*, 93(6), 445–458.
- Vlachos, J. (2012). Är vinst och konkurrens en bra modell för skolan? *Ekonomisk debatt*, nr. 4, 16–30.
- Wermke, W. (2013). *Development and autonomy: conceptualising teachers’ continuing professional development in different national contexts*. Doctoral dissertation. Stockholm: Stockholms universitet.
- Wigforss, F. (1933): Färdigheten i mekanisk räkning i folkskolan enligt Rostads standardtabeller. *Skolan och samhälle*, nr. 14, 191–222.
- Ydesen, C. (2011). *The Rise of High Stakes Educational Testing in Denmark, 1920–1970*. Frankfurt am Main: Peter Lang Verlag.