

Betygen i grundskolan – relationen mellan bedömning och elevers senare prestationer

ALLI KLAPP

Institutionen för pedagogik och specialpedagogik
Göteborgs universitet

Syftet med studien var att undersöka hur betyg påverkar elevers lärande mätt med betyg ett år senare, och hur betygsättning påverkar elever med olika förutsättningar utifrån kognitiv förmåga, kön och socioekonomisk bakgrund (Klapp, Cliffordson & Gustafsson, 2014). Data från databasen Utvärdering genom Uppföljning (UGU-projektet) har använts. På grund av att kommuner mellan åren 1969 och 1981 själva kunde bestämma om de skulle betygsätta elever i årskurs 6 har det varit möjligt att jämföra elever som fått betyg med elever som inte fått betyg i årskurs 6 och deras betyg i årskurs 7. Totalt deltog 8 558 elever födda 1967. Multipla regressionsanalyser genomfördes. Resultaten visade att det inte fanns några generella effekter av betygsättning på elevers prestationer ett år senare men det fanns differentierande effekter: betygsatta elever i årskurs 6 som presterade lågt på ett kognitiva test fick lägre betyg i årskurs 7 jämfört med elever som inte fick betyg i årskurs 6. Resultatet visar även interaktionseffekter mellan kön och betygsättning och kön och kognitiv förmåga.

INLEDNING

Under de senaste åren har flera reformer genomförts i skolan vars syfte har varit att skapa förutsättningar för att på ett mer systematiskt sätt mäta elevernas kunskaper och resultat. Betyg från årskurs 6, nationella prov i fler ämnen och i tidigare årskurser och en ny betygsskala har införts (Utbildningsdepartementet, 2013). I media har skoldiskussionen till stor del handlat om olika skolors prestationer på de nationella proven och hur Sverige rankas på de internationella kunskapsmätningarna som Program of International

Student Achievement (PISA) och the Trends in International Mathematics and Science Study (TIMSS). Både nationellt och globalt utvecklas och införs utvärderingssystem konstruerade på summativa bedömningar som tillmäts stor betydelse i diskussionen om kvalitet i skolan.

Samtidigt har forskare undersökt hur summativa bedömningar (prov och betyg) påverkar elevernas motivation till lärande och prestationer där resultaten visar att ett alltför stort fokus på resultat i skolan och användning av summativa bedömningar kan påverka elevernas lärande negativt (Black & Wiliam, 1998; Harlen & Deakin Crick, 2002; Hattie & Timperley, 2007; Sadler, 1998). Teorier som Self-Determination Theory (Deci, Koestner & Ryan, 2001; Cameron, 2001) och Conservation of Resources (COR) Theory (Covington, 2000; Frydenberg, 2008; Frydenberg & Lewis, 2009; Hobfoll, 1989, 2001) visar att det finns komplexa samband mellan stressfyllda situationer i skolan, som till exempel vid bedömning, självuppfattning och självkänsla, lärande och prestationer.

Bristen på storskaliga studier som undersökt effekter av summativa bedömningar på elevers lärande och prestationer beror rimligtvis på att det i de flesta utbildningssystem inte finns variation i betygsättningen. Antingen får alla elever i en viss årskurs betyg eller inte, vilket gör det svårt att empiriskt undersöka olika grupper av elever och jämföra hur de påverkas av betygsättning. Under en period på 12 år fanns det dock variation i betygsättning inom det svenska utbildningssystemet vilket har gjort det möjligt att jämföra betygsatta elever med icke betygsatta elever och deras senare prestationer.

Det övergripande syftet med studien är att undersöka hur betygsättning i årskurs 6 påverkar elevernas prestationer i årskurs 7. Ytterligare ett syfte är att undersöka betygens differentierande effekter för olika elevgrupper med avseende på kognitiv förmåga, kön och socioekonomisk bakgrund.

TIDIGARE FORSKNING

Hur påverkas elever av betyg?

I ett flertal studier har resultaten visat att återkoppling (feed-back) är ett effektivt verktyg för att utveckla elevernas lärande och prestationer i skolan (Hattie, 2009). Den positiva effekten av återkoppling utvecklas främst om den används i formativt syfte för att påverka både undervisningen och elevernas fortsatta lärande (Black & Wiliam, 1998; Deci m.fl., 2001; Hattie & Timperley, 2007). Harlen och Deakin Crick (2002) har i en översikt kring hur summativa bedömningar såsom prov och betyg påverkar elevers lärande, självkänsla och motivation dragit slutsatsen att lågpresterande elever påverkas negativt av summativa bedömningar jämfört med högpresterande elever

och att många elever upplever provångest. De menar även att elevers akademiska självkänsla kan påverkas negativt av summativa bedömningar och att de därför riskerar att utveckla distans och yttre motivation till sitt lärande. Flera andra forskare redovisar resultat i linje med detta, nämligen att summativa bedömningar påverkar elevernas prestationer och att olika grupper av elever påverkas olika, antingen mer negativt eller positivt, det vill säga att summativa bedömningar differentierar (Alexander, 2010; Brookhart, 1994; Natriello, 1987). Vidare menar Harlen och Deakin Crick (2002) att den existerande forskningen kring hur summativa bedömningar påverkar elevers motivation till lärande och prestationer visar på ett komplext mönster för olika grupper av elever och att teorier inom detta område behöver utvecklas.

Det har de senaste åren kommit resultat från två studier som visar på att betyg och information om rankning av elevers resultat påverkar de utsatta eleverna på ett positivt sätt (Azmat & Iriberri, 2009; Sjögren, 2010). Sjögren (2010) har i en studie jämfört betygsatta elever med icke betygsatta elevers utbildningslängd och lön vid 32 års ålder. Sjögren fann ingen generell effekt för pojkar, det vill säga att de inte verkar ha någon fördel av betygsättning för deras utbildningslängd och senare lön. Däremot fann Sjögren en signifikant generell effekt för flickor, där betygsatta flickor oavsett socioekonomisk bakgrund hade viss fördel av att få betyg i årskurs 6 jämfört med flickor som inte fått betyg i årskurs 6. Denna effekt var tydligast för lågpresterande flickor. Lågpresterande pojkar med låg socioekonomisk bakgrund verkade också ha viss fördel av att få tidiga betyg.

Azmat och Iriberri (2009) undersökte elever i en privat skola på gymnasienivå och hur deras senare prestationer påverkades av att de under ett läsår fick information om prestationer för övriga elever i klassen. Eleverna fick betyg fyra gånger under ett läsår samt information om hur de rankades i klassen och eleverna kunde därför jämföra sitt eget resultat med ett genomsnitt för klassen. Azmat och Iriberri (2009) fann att främst de låg- och högpresterande eleverna hade fördel av att få extra information om rankning. Azmat och Iriberri menar att den positiva effekten berodde på att eleverna fick ny information som gjorde det möjligt till jämförelse och att det ledde till bättre prestationer. De drar slutsatsen att information om rankning underlättar sociala jämförelser och att det är positivt för elevernas senare prestationer, speciellt för de låg- och högpresterande eleverna.

Betydelsen av elevernas individuella förutsättningar

Flera studier har visat att elevers olika förutsättningar och bakgrund har betydelse för hur de påverkas av summativa bedömningar (Harlen & Deakin Crick, 2002; Natriello, 1987). Resursstarka och högpresterande elever verkar ha en bättre förståelse av betygens innebörd och blir mindre negativt påver-

kade av den återkoppling som lärare ger dem jämfört med mer resurssvaga och lågpresterande elever (Harlen & Deakin Crick, 2002; Butler, 1988). Butler (1988) undersökte och jämförde elever i årskurs 5 och 6 där en grupp elever fick återkoppling på uppgifter i skolan genom enbart poäng, en andra grupp fick enbart kommentarer medan en tredje grupp fick både poäng och kommentarer. Resultatet visade att både låg- och högpresterande elever behöll sitt intresse för uppgifterna och kände starkare motivation för uppgifterna om de enbart fick kommentarer som återkoppling. När återkopplingen bestod av poäng blev både låg- och högpresterande elever negativt påverkade med avseende på intresse och prestation, dock behöll de högpresterande eleverna intresse för arbetet i större utsträckning när både kommentarer och poäng gavs jämfört med de lågpresterande eleverna.

Skillnader i betyg mellan pojkar och flickor ökar till flickors fördel. Pojkar brukar ha mer positiva självvärderingar inom områden som matematik, naturkunskap och gymnastik medan flickor tenderar att värdera sig själva som duktiga inom språk, läsning och konst (Gustafsson & Yang, 2009). Voyer och Voyer (2014) har i en metaanalys undersökt könsskillnader i betyg mellan åren 1911-2013 och fann att flickornas högre betyg inte är ett nytt fenomen utan har funnits under lång tid. De fann att flickors fördel i betygen gäller inom alla ämnesdomäner.

Flickor säger sig även i större utsträckning uppleva nervositet inför prov liksom att de förklarar svårigheter och misslyckanden med skolarbetet utifrån sina egna personliga och inre attribut vilket kan leda till lägre självkänsla jämfört med pojkarna som förklarar misslyckanden utifrån yttre attribut (Johnston & McClune, 2000).

Evans och Engelberg (1988) undersökte innebörden av betyg för elever i olika åldrar och menar att äldre elever förstår innebörden av betyg bättre jämfört med yngre elever men att även äldre elever kan ha svårigheter med att förstå komplexa betygssystem. Att få ett betyg utan att förstå innebörden av betyg kan leda till att elever känner hjälplöshet (Evans & Engelberg, 1988). Studier som undersökt effekter av betyg för elever med olika socioekonomiska bakgrunder är dock få (Harlen & Deakin Crick, 2002).

Relationen mellan bedömning, självkänsla och prestationer

Inom flera motivationsteorier lyfts de medierande relationerna mellan bedömning, självkänsla och prestationer fram för att förklara elevers framgång och misslyckande i skolan och senare i livet (Bandura, 1986; Deci m.fl., 2001; Frydenberg, 2008; Hobfoll, 1989, 2001). Conservation of Resources Theory (COR) (Hobfoll, 1989, 2001; Frydenberg, 2008) är en övergripande stress- och motivationsteori som diskuterar hur olika personliga resurser och tillgång till dessa påverkar människors liv, till exempel hur elever påverkas

av olika stressfyllda situationer i skolan. Resurser är personliga egenskaper, förutsättningar och tillstånd som självkänsla, attityder, social kompetens, uppfattning om den egna kompetensen och att kunna reglera egna känslor. Elever strävar efter att behålla, få, skydda och utveckla de egna resurserna som värderas högt inom en grupp, till exempel i skolan för att minimera psykisk stress. När elever erfar förluster av resurser ökar risken för ytliga lärandestrategier och negativa attityder till skolan. Vid förluster av resurser riskerar eleven att känna frustration vilket i sin tur kan orsaka externaliserande och internaliserande beteendeproblem (Hobfoll, 2001; Frydenberg, 2008).

En princip inom COR-teorin handlar om den asymmetriska relationen mellan förlust och erhållande av resurser och hur människor påverkas. Förlust av en resurs som till exempel självkänsla är disproportionellt mer negativt jämfört med att erhålla, få eller utveckla resurser. Om en elev upplever lika mycket resursförlust som att få en resurs, påverkar resursförluster eleven mer. Flera forskare har studerat detta och denna ”negative bias” innebär att människor tillskriver negativ information större betydelse jämfört med positiv information (Cacioppo & Gardner, 1999; Hobfoll, 2001; Taylor, 1991).

Sammanfattningsvis visar forskningen inom detta område på en konsensus att summativa bedömningar har differentierande effekter. Dock är resultaten till viss del disparata om hur betyg differentierar och hur olika elevgrupper påverkas av betygsättning.

Det svenska bedömningssystemet mellan åren 1969 och 1981

Under 12 år, mellan åren 1969 till 1981, var det frivilligt för kommunerna i Sverige att betygsätta elever i årskurs 3 och 6. Detta innebar att i vissa kommuner övergavs betygen i årskurs 3 och 6 i början av tidsperioden medan andra kommuner valde att behålla betygen i både årskurs 3 och 6 ända fram till 1981. År 1982 var dock alla kommuner tvungna att upphöra med betygen före årskurs 7. Detta ”naturliga experiment” har gjort det möjligt att i en storskalig studie jämföra elever som fått betyg i årskurs 6 med elever som inte fått betyg i årskurs 6 och hur de presterar i årskurs 7, samt hur betyg påverkar olika grupper av elever utifrån kognitiv förmåga, kön och socioekonomisk bakgrund (Klapp m.fl., 2014).

DESIGN AV STUDIEN

Studien har dragit fördel av att deltagarna har följt samma läroplan och läst samma ämnen genom grundskolan. Alla elever fick betyg i årskurs 7. Däremot fick cirka 50 procent av eleverna betyg i årskurs 6 medan 50 procent inte fick betyg i årskurs 6. Denna omständighet har gjort det möjligt att

undersöka hur betygsatta elever i årskurs 6 presterar ett år senare (genom betygen i årskurs 7). Alla elever har läst samma ämnen (från årskurs 1 eller 3), ämnen som eleverna har fått undervisning i genom grundskolan: svenska, matematik, engelska, orienteringsämne, slöjd, teckning, musik och gymnastik (pågående ämnen). I årskurs 7 introducerades tre nya ämnen: biologi, fysik och kemi (nya ämnen). Detta har gjort det möjligt att jämföra betygen i årskurs 7 för pågående och nya ämnen. De betygsatta eleverna har fått betyg i årskurs 6 i de pågående ämnen men inte fått betyg i de tre nya ämnena i årskurs 7. Betygen i de pågående ämnena är påverkade av flera tidigare bedömningar och betyg.

Syfte

Det övergripande syftet med studien är att undersöka hur betygsättning i årskurs 6 påverkar elevernas prestationer i årskurs 7. Ytterligare ett syfte är att undersöka betygens differentierande effekter för olika elevgrupper utifrån kognitiv förmåga, kön och socioekonomisk bakgrund.

Hypotesen är att betygsättning i årskurs 6 kommer att påverka betygen i de pågående ämnena i årskurs 7 negativt medan det inte kommer att finnas någon effekt eller enbart en svag effekt av betygsättning i årskurs 6 för betygen i de nya ämnena i årskurs 7.

Metod och data

Data hämtades från databasen Utvärdering genom Uppföljning (UGU) och The Gothenburg Longitudinal Database (GOLD) där det finns information om elever som under dessa år fått betyg eller inte samt en mängd information om elevernas bakgrund och senare prestationer (se Härnqvist, 2000). UGU-databasen innehåller data för ett nationellt representativt urval om 10 procent av en kohort elever i Sverige (cirka 10 000 elever) för ett flertal kohorter samt både enkät- och registerdata. Totalt 8 558 elever födda 1967 och som gick ut grundskolan 1983 deltog i studien. I studien har information om betyg i årskurs 7, resultat från ett kognitivt test i årskurs 6, kön och socioekonomisk bakgrund (SES) använts.

Betygsättning

Läroplanen som trädde i kraft 1969 tillät variation på kommunal nivå huruvida elever skulle betygsättas eller inte i årskurs 6. Från 1982 övergavs dock betygsättning innan årskurs 7 för alla kommuner. Skillnader mellan kommuner (kommunkaraktäristika) och tidpunkt för när betyg i årskurs 6 upphörde i de olika kommunerna skulle kunna vara en orsak till snedvridning av resultaten. Detta hanterades genom det kognitiva testet i årskurs 6 som

speglar elevernas kognitiva förmåga vilket innebär att det går att jämföra de två grupperna av elever (betygsatta eller inte). Tillgången till information om elevers kognitiva förmåga, socioekonomisk bakgrund och kön gör det möjligt att kontrollera för skillnader mellan betygsatta och inte betygsatta elever.

Hälften av eleverna fick betyg i årskurs 6, totalt 4 151 elever (48,5 %) medan 4 407 elever (51,5 %) inte fick betyg. Betygsättning är en dikotom variabel där elever som inte fick betyg i årskurs 6 är kodade som 0 medan elever som fick betyg i årskurs 6 är kodade som 1.

Kognitivt test, kön och socioekonomisk bakgrund

För att ha en utgångspunkt för elevernas förmåga användes ett kognitivt test (KOGN) som har tre delar: induktivt, spatialt och verbalt. Det verbala testet mäter vokabulär genom antonymer, det spatiala testet mäter förmåga att visualisera genom plåtvikningsuppgifter medan det induktiva testet mäter numeriskt tänkande genom nummerserier. Varje test innehöll 40 uppgifter. Alla eleverna i studien genomförde testen på våren i årskurs 6 och resultat från de tre deltesten summerades och omkodades till standardiserade värden (-3,49 till 2,64). De kognitiva testen har visats vara ett tillförlitligt mått på elevernas kognitiva förmåga utan större skillnader mellan elever beroende på socioekonomiska bakgrund och kön (Svensson, 1971). Det kognitiva testet mäter generella kognitiva förmågor och är inte ett prov som är relaterat till kunskaper i de olika skolämnena. Korrelationer mellan de olika delarna i det kognitiva testet och prestationer i skolan (betyg) sträcker sig från 0,58 till 0,63 för eleverna i studien. I studien används det kognitiva testet som en kontrollvariabel för elevernas generella kognitiva förmåga, som kan anses vara relativt opåverkat av bakgrundsfaktorer och som inte primärt mäter specifika skol- och ämneskunskaper utvecklade i skolan.

Information om elevernas socioekonomiska bakgrund (SES) är ett index konstruerat av föräldrars inkomst, utbildningsnivå och yrke och kodat SES III = 0, SES II = 1, SES I = 2 och används som en kontinuerlig variabel i analyserna. Kön är en dikotom variabel där pojke = 0 och flicka = 1.

Betyg och betygsgenomsnitt i årskurs 7

Betygen som användes då data samlades in var normrelaterade och sträckte sig från 1 till 5, där 5 var det högsta betyget. I årskurs 7 fick alla elever i studien betyg i 14 ämnen varav tre ämnen introducerades för första gången i årskurs 7: biologi, fysik och kemi. Designen av studien drar fördel av att de betygsatta eleverna fått undervisning och bedömts i de pågående ämnena (PÅG_ÄM) i årskurs 6 medan inga elever fått tidigare bedömning i de nya ämnena (NYA_ÄM). PÅG_ÄM innehåller 11 betyg i följande ämnen:

svenska, engelska, matematik, historia, samhällskunskap, geografi, religion, musik, slöjd, teckning och gymnastik. NYA_ÄM innehåller betyg i tre ämnen: biologi, fysik och kemi.

Analysmetod

Ett flertal multivariata regressionsanalyser genomfördes i Mplus (Muthén & Muthén, 2009). Först estimerades en basmodell för att undersöka huvudeffekterna av BETYG, KOGN, KÖN, SES för PÅG_ÄM och NYA_ÄM med kovarians mellan oberoende variabler och residualerna för de beroende variablerna. I nästa steg estimerades en full modell med både huvudeffekter av BETYG, KOGN, KÖN och SES och första-, andra- och tredjehands interaktionseffekter. I den sista modellen estimerades enbart de signifikanta relationerna i de tidigare modellerna. För en mer utförlig beskrivning av analyser, modellering och resultat, se Klapp med flera (2014).

Missing data modellering har använts i studien, vilket innebär att uteblivna svar som bildar ett gemensamt mönster modelleras i en egen matris. Sedan kombineras de olika missing-datamatriskerna till en total matris (Muthén, Kaplan & Hollis, 1987).

Klustrade data är vanligt i utbildningssammanhang och innebär att elever på en skola har mer gemensamt med varandra jämfört med elever på andra skolor. För att ta hänsyn till att data "klustrar" sig har complex-funktionen i Mplus använts vilken tar hänsyn till klustrade data och justerar anpassningsmåttan medan estimaten förblir desamma (Muthén & Muthén, 2009).

Effektstorlekar (Cohen's *d*) redovisas för delar av resultatet och visar skillnader mellan betygsatta och icke betygsatta elever samt mellan pojkar och flickor. En effektstorlek på över 0,10 är värd att ta i beaktning när urvalet är stort och multipla variabler används (Ellis, 2010).

RESULTAT

I tabell 1 redovisas deskriptiv statistik för PÅG_ÄM och NYA_ÄM i årskurs 7 och KOGN i årskurs 6, uppdelat på betygsatta och icke betygsatta elever samt kön. Missing data för PÅG_ÄM och NYA_ÄM samt KOGN var mellan 7,9 och 10,5 procent.

Tabell 1. Deskriptiv statistik för pågående ämnen (PÅG_ÄM), nya ämnen (NYA_ÄM) och kognitiv förmåga (KOGN) för alla elever (Total), pojkar och flickor och uppdelade med avseende på elever som inte fått betyg och elever som fått betyg.

| Variabler | Inga Betyg = 0 | | | | | | Betyg = 1 | | | | | |
|-------------------|----------------|------|---------|------|-------|------|-----------|------|---------|------|-------|------|
| | Pojkar | | Flickor | | Total | | Pojkar | | Flickor | | Total | |
| | N | % | N | % | N | % | N | % | N | % | N | % |
| | 2 183 | 25,5 | 2 224 | 26,0 | 4 407 | 51,5 | 2 144 | 25,0 | 2 007 | 23,5 | 4 151 | 48,5 |
| | M | Sd | M | Sd | M | Sd | M | Sd | M | Sd | M | Sd |
| PÅG_ÄM | 3,07 | 0,61 | 3,29 | 0,59 | 3,18 | 0,61 | 3,02 | 0,67 | 3,35 | 0,62 | 3,18 | 0,66 |
| NYA_ÄM | 2,97 | 0,78 | 3,08 | 0,75 | 3,02 | 0,77 | 2,95 | 0,82 | 3,16 | 0,74 | 3,05 | 0,79 |
| ¹ KOGN | -0,03 | 1,00 | -0,03 | 0,99 | -0,03 | 1,00 | 0,00 | 1,03 | 0,07 | 0,97 | 0,03 | 1,00 |

Not: ¹KOGN= Summan av de tre kognitiva testen, standardiserade till z-poäng (från -3,49 till 2,64).

Oberoende t-test genomfördes för att undersöka skillnader mellan grupper av elever (BETYG och KÖN) för PÅG_ÄM, NYA_ÄM, KOGN och SES. Resultatet visade inga signifikanta skillnader med avseende på BETYG för PÅG_ÄM, NYA_ÄM, KOGN och SES. Resultatet visade dock signifikanta skillnader med avseende på KÖN för både PÅG_ÄM $t(7878) = 10,23$, $p = 0,00$ och NYA_ÄM $t(7678) = 3,42$, $p = 0,00$. Flickorna hade högre betyg i PÅG_ÄM och NYA_ÄM jämfört med pojkarna (Klapp m.fl., 2014).

Det inledande resultatet (modell 1 i tabell 2) visade inga signifikanta huvudeffekter av BETYG för PÅG_ÄM eller NYA_ÄM medan kraftfulla huvudeffekter existerade för KOGN, KÖN och SES. Detta innebär att elever med högt resultat på det kognitiva testet, flickor samt elever med hög socioekonomisk bakgrund fick högre betyg i årskurs 7 jämfört med elever som hade motsatta karaktärstika.

När interaktionseffekter inkluderades i modellen (modell 2 i tabell 2) blev BETYG signifikant för PÅG_ÄM men icke signifikant för NYA_ÄM, vilket var i linje med hypotesen att effekter av BETYG främst skulle finnas i PÅG_ÄM. Resultatet visade på signifikanta interaktionseffekter för BETYG*KOGN, BETYG*KÖN samt KÖN*KOGN som därför inkluderades i den sista modellen (modell 3 i tabell 2).

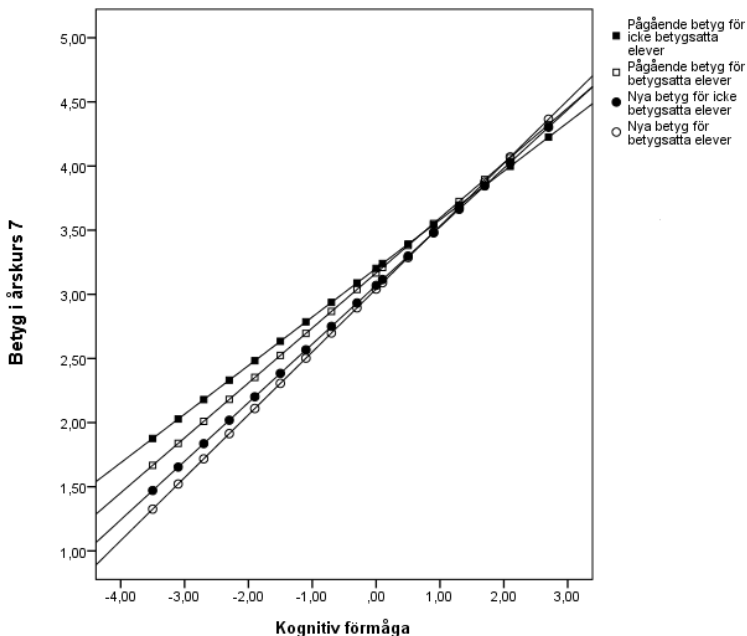
Tabell 2. Standardiserade regressionskoefficienter och t-värden för de tre multipla regressionsmodeller med PÅG_ÄM och NYA_ÄM som beroende variabler, kovarians mellan de oberoende variablerna och mellan residualerna för de beroende variablerna.

| Variabler | Modell 1 | | | Modell 2 | | | Modell 3 | | | | | |
|--------------------|----------|--------|----------|----------|----------|---------|----------|---------|----------|----------|----------|-------|
| | PÅG_ÄM | NYA_ÄM | | PÅG_ÄM | NYA_ÄM | | PÅG_ÄM | NYA_ÄM | | | | |
| | β | t | | β | t | | β | t | | | | |
| ¹ BETYG | -0,010 | -0,79 | -0,003 | -0,024 | -0,045** | -2,84 | -0,032 | -1,81 | -0,041** | -2,68 | -0,028 | -1,62 |
| ² KOGN | 0,597*** | 72,27 | 0,579*** | 64,57 | 0,560*** | 28,07 | 0,583*** | 27,60 | 0,571*** | 33,12 | 0,582*** | 31,76 |
| ³ KÖN | 0,202*** | 20,02 | 0,089*** | 7,63 | 0,170*** | 13,05 | 0,058*** | 3,57 | 0,168*** | 13,00 | 0,062*** | 3,93 |
| ⁴ SES | 0,134*** | 13,99 | 0,121*** | 12,18 | 0,148*** | 7,65 | 0,126*** | 7,22 | 0,140*** | 14,69 | 0,125*** | 12,66 |
| BETYG*KOGN | | | | 0,047* | 2,39 | 0,018 | 0,82 | 0,041** | 2,84 | 0,022 | 1,41 | |
| BETYG*KÖN | | | | 0,056** | 3,13 | 0,047* | 2,26 | 0,058** | 3,28 | 0,047* | 2,37 | |
| KÖN*KOGN | | | | -0,009 | -0,47 | -0,043* | -2,21 | -0,020 | -1,52 | -0,042** | -3,02 | |
| BETYG*SES | | | | -0,009 | -0,47 | 0,002 | 0,12 | | | | | |
| KÖN*SES | | | | -0,026 | -1,51 | -0,012 | -0,66 | | | | | |
| KOGN*SES | | | | -0,013 | -0,66 | -0,023 | -1,10 | | | | | |
| BETYG*KOGN*KÖN | | | | -0,009 | -0,49 | 0,008 | 0,40 | | | | | |
| BETYG*KOGN*SES | | | | 0,015 | 0,75 | 0,017 | 0,85 | | | | | |
| KÖN*KOGN*SES | | | | -0,003 | -0,16 | 0,018 | 1,05 | | | | | |
| BETYG*KÖN*SES | | | | 0,019 | 1,04 | -0,003 | -0,16 | | | | | |
| BETYG*KOGN* | | | | 0,007 | 0,41 | 0,003 | 0,17 | | | | | |
| KÖN*SES | | | | | | | | | | | | |
| R ² | 0,46 | | 0,39 | | 0,45 | | 0,38 | | 0,45 | | 0,39 | |

¹Inte fått betyg = 0 och fått betyg = 1; ²z-score; ³pojkar = 0 och flickor = 1; ⁴SES kodad: SES III = 0, SES II = 1 and SES I = 2. * p<0,05; ** p<0,01; *** p<0,001.

Resultatet från den sista modellen (modell 3) visade att elever som presterade lågt på det kognitiva testet fick lägre betyg i årskurs 7 om de hade fått betyg i årskurs 6 jämfört med elever som presterade lågt på det kognitiva testet och som inte fått betyg i årskurs 6, se figur 1. Resultatet stödjer hypotesen att de negativa effekterna av BETYGG skulle finnas i PÅG_ÄM eftersom eleverna fått tidigare bedömningar i de ämnena. Inga effekter, varken negativa eller positiva, av BETYGG fanns i NYA_ÄM. Den svagt positiva effekten för betygsatta elever som presterade högt på det kognitiva testet var inte signifikant.

Resultatet visade på ett liknande mönster för hur BETYGG påverkade pojkar och flickor men det fanns en förskjutning mot att pojkar fick lägre betyg i årskurs 7 om de hade fått betyg i årskurs 6, jämfört med flickor. Effektstorleken (Cohen's d) för skillnaderna mellan betygsatta och inte betygsatta elever sträcker sig från $d = -0,30$ till $d = 0,14$ för PÅG_ÄM (från den lägsta gemensamma poängen på det kognitiva testet till den högsta gemensamma poängen på det kognitiva testet). För betygen i de NYA_ÄM var effektstorlekarna lägre (från $d = -0,17$ till $d = 0,07$). Skillnaderna mellan betygsatta och icke betygsatta elever var störst för de elever som presterade lågt på det kognitiva testet.



Figur 1. Elever som presterade lågt på det kognitiva testet fick lägre betyg i årskurs 7 om de hade fått betyg i årskurs 6 jämfört med elever som presterade lågt på det kognitiva testet och som inte fått betyg i årskurs 6.

Interaktionseffekten BETYG*KÖN visade att det finns en starkare könsskillnad mellan betygsatta elever jämfört med icke betygsatta elever. Skillnaden var störst mellan betygsatta pojkar och flickor för PÅG_ÄM, Cohen's $d = 0,51$, till flickors fördel. För NYA_ÄM visade resultatet på liknande mönster som för PÅG_ÄM men avsevärt svagare.

Interaktionseffekten KÖN*KOGN visade på vissa skillnader för PÅG_ÄM och NYA_ÄM. Resultatet visade att oavsett kognitiv förmåga fick pojkar lägre betyg jämfört med flickor i PÅG_ÄM. Flickor som presterade lågt på det kognitiva testet fick högre betyg i NYA_ÄM jämfört med pojkar som presterade lågt på det kognitiva testet. Dock fanns inga könsskillnader för elever som presterade högt på det kognitiva testet i de nya ämnena. För en mer ingående resultatredovisning, se Klapp med flera (2014).

Sammanfattningsvis visar resultaten att det finns differentierande effekter av betygsättning på elevernas senare prestationer. Elever som presterade ett lågt till medelresultat på det kognitiva testet och som blev betygsatta i årskurs 6 fick lägre betyg i årskurs 7 jämfört med icke betygsatta elever som presterade lågt på det kognitiva testet med kontroll för kön och socioekonomisk bakgrund. Det finns skillnader mellan låg- och högpresterande (på det kognitiva testet) elever och mellan pojkar och flickor när det gäller i vilken grad de påverkas av summativa bedömningar som betyg. Däremot fanns inga skillnader mellan elevernas socioekonomiska bakgrund och hur de påverkades av betygsättning.

DISKUSSION AV STUDIENS RESULTAT

Syftet med studien var att undersöka hur betygsättning i årskurs 6 påverkade elevernas betyg i årskurs 7 och hur effekten av betygsättning påverkar olika grupper av elever med avseende på kognitiv förmåga, kön och socioekonomisk bakgrund. Resultatet visade inga huvudeffekter av betygsättning på senare prestationer och betyg. Däremot fanns differentierande effekter där betygsatta elever som presterade lågt på det kognitiva testet (resurssvaga elever) fick lägre betyg i årskurs 7 jämför med elever med liknande förutsättningar (resurssvaga elever) som inte fick betyg i årskurs 6. Resultatet visade på en tendens till positiva effekter av betygsättning för elever som presterade högt på det kognitiva testet men denna tendens var svag och icke signifikant och stödjer inte slutsatsen att betyg påverkar resursstarka elever positivt. Resultatet visade även på könsskillnader där pojkar får lägre betyg jämfört med flickor, givet samma kognitiva förmåga (Klapp m.fl., 2014). Pojkar är generellt mer negativt påverkade av betygsättning jämfört med flickor och könsskillnaderna är större inom gruppen betygsatta elever, jämfört med gruppen inte betygsatta elever.

När eleverna fick betyg i årskurs 6 fick de ny information om sin kunskapsutveckling och den nya informationen kan ha varit mer eller mindre i linje med elevens egen uppfattning om sin kunskapsutveckling. Om en elev får ett högre betyg är förväntat är det rimligt att anta att betygen fungerar som en belöning medan det omvända kan antas råda för en elev som får lägre betyg är förväntat (Black & Wiliam, 1998; Deci m.fl., 2001; Hattie, 2009).

En möjlig förklaring till de negativa effekterna av betyg för resurssvaga elever är att ny information som är negativ (ett lågt betyg) innebär en förlust av självförtroende vilket i sin tur kan leda till stress och ytliga lärandestrategier och negativ attityd till skolan vilket i sin tur kan påverka senare prestationer mätt i betyg (Covington, 2000; Frydenberg, 2008; Hobfoll, 2001). Den asymmetriska relationen mellan negativ och positiv återkoppling som Taylor (1991) diskuterar kan förklara varför det finns en mer kraftfull negativ effekten av betygsättning för de resurssvaga eleverna jämfört med den svaga positiva effekten av betygsättning för de resurstarka eleverna. En negativ återkoppling, som ett lägre betyg än förväntat, verkar kunna ha en starkare negativ påverkan på senare prestationer, jämfört med effekten av en positiv återkoppling för senare prestationer (Klapp m.fl., 2014).

Studiens resultat i relation till tidigare forskning

Resultaten som diskuteras i denna artikel visar på att betygen har differentierande effekter, att elever påverkas på olika sätt av betyg beroende på deras förmåga och kön. Resultaten är till största del i linje med tidigare internationell forskning som har visat att betyg påverkar lågpresterande elever negativt. Eftersom Sjögrens (2010) och Azmat och Iriberris (2009) resultat står i viss motsättning till internationell forskning samt resultatet i denna studie är det intressant att diskutera möjliga orsaker till dessa skillnader.

Sjögren (2010) använde samma reform men använde annan data i sin studie och undersökte betygsatta och inte betygsatta elevers utbildningslängd och lön vid 32 års ålder. En möjlig orsak till skillnaderna kan vara att det i Sjögrens studie fanns effekter på kommunnivå som påverkar resultatet, till exempel att politiskt styre och kommuninvånarnas socioekonomiska tillhörighet påverkade resultatet. Sjögren har dock kontrollerat för kommuneffekter och även om där finns sådana verkar de vara för svaga för att ha kunnat påverka resultatet. Sjögren använder ett indirekt mått på elevernas förmåga (om de avgått gymnasiet eller inte) vilket kan vara ett osäkert mått på förmåga. Att dra slutsatser huruvida betygsättning i årskurs 6 påverkar elevernas utbildningslängd och lön vid 32 års ålder kräver att kontroll för en mängd andra mekanismer och faktorer görs eftersom de kan ha påverkat utbildningslängd och lön.

Azmat och Iriberry (2009) fann att om information om rankning gavs till eleverna presterade de bättre, vilket inte är i linje med resultaten funna i denna studie. En möjlig förklaring är att Azmat och Iriberry (2009) undersökte en selekterad grupp elever på en privat skola på gymnasienivå som förmodligen är mer homogen när det gäller förmåga, jämfört med elever i den svenska grundskolan. Den information om rankning som gavs var troligen inte tydlig när det gällde skillnaden mellan elevernas uppfattning av sin egen kompetens och sina akademiska prestationer utan gav snarare information om hur de andra eleverna i klassen ansträngde sig. Detta kan ha påverkat eleverna att anstränga sig mer och då kan en generell positiv effekt av rankning ha uppstått. En möjlig förklaring är att gruppen elever som undersöktes var resursstarka vilket kan innebära en generell positiv effekt på elevernas prestationer när information om rankning gavs.

Studiens resultat i relation till relevanta teorier

Både internationell och svensk forskning visar att betyg påverkar elever på olika sätt och att det finns skillnader med avseende på kön och kognitiv förmåga (Klapp m.fl., 2014; Sjögren, 2010). Om eleven får positiv återkoppling som är i linje med vad eleven förväntar sig verkar det troligt att elevens självkänsla stärks (Frydenberg, 2008; Frydenberg & Lewis, 2009; Hobfoll, 2001). Om en elev däremot får negativ återkoppling som inte är i linje med vad eleven förväntat sig kan en negativ lärande- och prestationsloop startas. När eleverna blir betygsatta får de en bedömning som kan påverka deras syn på den egna kompetensen och deras möjligheter att lära sig samt om de kan lyckas inom utbildningssystemet. Elever behöver utveckla en positiv självbild och en positiv uppfattning av den egna akademiska kompetensen eftersom de tillbringar en stor del av sin tid i skolan där dessa kompetenser värderas positivt och är viktiga för framgång.

Resultatet från studien visar att de negativa effekterna av betygsättning på senare prestationer är kraftfullare än de positiva effekterna. Den mer kraftfulla negativa effekten av betygsättning kan möjligen förklaras av hur elever hanterar förluster av resurser, till exempel dåliga resultat i skolan där självkänslan påverkas negativt. Detta kan innebära att eleven utvecklar defensiva strategier som förnekelse och ”konserverar” sina resurser (Hobfoll, 2001) och en accelererande negativ spiral kan skapas (Taylor, 1991). Eftersom den asymmetriska relationen mellan negativ och positiv information (att förluster har kraftfullare negativ påverkan på individer jämfört med att få resurser) har negativa förlust-spiraler en kraftfullare påverkan på eleverna och accelererar i högre hastighet.

De elever som däremot har tillgång till högt värderade resurser kan vid misslyckande i skolan använda förnekelse och icke produktiva strategier under en begränsad tid och återfår produktiva strategier snabbare (Frydenberg & Lewis, 2009). I undervisningssammanhang innebär detta att elever bör lyckas i så stor utsträckning som möjligt eftersom det stärker och utvecklar elevens resurser och minskar risken för förluster av resurser. Att förstå hur elever påverkas av stressfyllda situationer innebär att det är yttre omständigheter, situationer och händelser som behöver fokuseras. COR-teorin kan användas som en förklaringsmodell för hur gemensamma kulturella tolkningar av problem leder till stress hos individer. Att förhindra och ta bort problematiska hinder i "miljön" för att nå en bättre anpassning till individernas resurser i en viss situation är centralt för COR-teorin, till exempel att elever oavsett kön eller social tillhörighet har tillgång till högre utbildning.

Betyg är rimligtvis viktiga för eleverna. Oavsett om eleverna tar skolan på allvar eller inte, om de är låg- eller högpresterande, pojkar eller flickor, blir de förmodligen påverkade av betyg och andra resultat i skolan. Det är rimligt att tro att resurssvaga och lågpresterande elever till större del har erfart misslyckanden i skolan, negativa bedömningar och har negativa erfarenheter med sig från skolan, deras syn på sig själva, sina kunskaper och kompetenser, jämfört med resursstarka och högpresterande elever. Att få betyg visar på ett konkret sätt hur det har gått i skolan och om elevernas och föräldrarnas förväntningar har infriats. Sammanfattningsvis finns det skäl att tro att de komplexa sambanden mellan elevernas självkänsla och uppfattning om sin egen kompetens, hur de upplever stressfyllda situationer i skolan som till exempel vid bedömning och betygsättning och hur de lyckas senare i livet, har betydelse för att förklara varför betyg differentierar.

Avslutande diskussion

Att betygsättning i sig självt är negativt är förmodligen en förenkling på grund av att andra faktorer utelämnats som förklaringsmodell till varför resurssvaga elever blir negativt påverkade av betyg. Utelämnade faktorer kan vara graden av differentiering i undervisningen och bedömningskulturen i klassrummet och om och hur elever får möjlighet att utveckla sina resurser som till exempel självkänsla och uppfattning av den egna akademiska kompetensen. Elevers självkänsla är förmodligen en central aspekt för att förstå och förklara betygens differentierande effekter (Hobfoll, 1989, 2001).

I debatten kring de försämrade resultaten från de senaste PISA undersökningarna lyfts bland annat ökad utvärdering, kontroll och betyg fram som ett sätt att nå förändring och bryta den nedåtgående trenden i skolans resultat.

Betygen ges då en informrande och en kontrollerande funktion med avsikt att den information som betygen ger kan förändra elevernas prestationer.

Ett gemensamt mål för många länder är att eleverna ska lyckas och att så många som möjligt når framgång inom skola och arbetsliv. De statliga besparingarna är mycket stora (Levin, 2009) och de personliga vinsterna likaså vid framgång inom skola och arbetsliv. Forskning visar också att elever med sämre förutsättningar påverkas mer negativt av differentierande skol- och bedömningssystem och att denna typ av system ökar gapet mellan låg- och högpresterande elever (Harlen & Deakin Crick, 2002). Andel elever i Sverige som inte kan påbörja gymnasiala studier ökar och var 13,1 procent våren 2014 (Skolverket, 2014). Att gapet är litet mellan låg- och högpresterande elever liksom att skillnaderna är små mellan människors livsvillkor tycks vara en viktig framgångsfaktor för samhället i stort. Sveriges nuvarande bedömnings- och skolsystem har utvecklats från att tidigare (fram till början av 1990-talet) ha varit ett lågt differentierat till att numera vara ett högt differentierat skol- och bedömningssystem vilket riskerar att föra med sig ökad segregation och att gapet mellan resurssvaga och resursstarka elever ökar (Skolverket, 2014). Konsekvenserna av ett högt differentierande utbildningssystem bör i större utsträckning analyseras och diskuteras i relation till internationella och nationella forskningsresultat inom området.

Begränsningar

En möjlig invändning mot resultaten i den nyligen genomförda studien (Klapp m.fl., 2014) kan vara om det är rimligt att använda data som är cirka 30 år gamla och om det går att generalisera resultaten till det nuvarande betygssystemet. När data samlades in hade Sverige ett normrelaterat betygssystem med en skala som gick från 1 till 5. Elever i detta system kunde inte få underkända betyg och var därför behöriga till gymnasiestudier även med de lägsta betygen. I dagens kriterierelaterade betygssystem finns en underkänt-nivå (F) som gör att elever riskerar att inte få behörighet till gymnasiala studier. Effekten av ett underkänt betyg i det nuvarande systemet kan därför anses vara större jämfört med det gamla betygssystemet, vilket gör att betygen idag är av större vikt för elevers möjlighet att kunna läsa vidare. Detta gör att de negativa effekterna av betygsättning på prestationer i skolan kan förväntas vara större för elever i dagens betygssystem.

En annan begränsning i studien är att elevernas självkänsla och syn på sin egen akademiska kompetens inte analyserats i relation till hur de påverkas av betyg. I en pågående studie analyseras enkätadata, som har information om elevers självkänsla och uppfattningar om kompetens, i relation till betygens differentierande effekter.

REFERENSER

- Alexander, R. (Ed.) (2010). *Children, their World, their Education. Final report and recommendations of the Cambridge Primary Review*. London: Routledge.
- Azmat, G. & Iriberry, N. (2009). *The importance of relative performance feedback information: Evidence from a natural experiment using high school students*. CEP Discussion Paper No 915, Centre for Economic Performance, London School of Economics and Political Science.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
- Brookhart, S. M. (1994). Teachers' grading: practice and theory. *Applied Measurement in Education*, 7(4), 279–301.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58(1), 1–14.
- Cacioppo, J. T. & Gardner, W. L. (1999). Emotion. *Annual Review of Psychology*, 50, 191–214.
- Cameron, J. (2001). Negative effects of reward on intrinsic motivation – a limited phenomenon: Comment on Deci, Koestner, and Ryan (2001). *Review of Educational Research*, 71(1), 29–42.
- Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, 51, 171–200.
- Deci, E. L., Koestner, R. & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, 71(1), 1–27.
- Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analyses and the interpretation of research results*. New York, NY: Cambridge University Press.
- Evans, E. & Engelberg, R. (1988). Pupils' perceptions of school grading, *Journal of Research and Development in Education*, 21, 44–54.
- Frydenberg, E. (2008). *Adolescent coping: Advances in theory, research and practice*. London & New York, NY: Routledge, Taylor & Francis Group.

- Frydenberg, E. & Lewis, C. (2009). The relationship between problem-solving efficacy and coping among Australian adolescents. *British Journal of Guidance and Counselling*, 37(1), 51–64.
- Gustafsson, J.-E. & Yang Hansen, K. (2009). Resultatutvecklingen i den svenska grundskolan. I Skolverket (2009), *Vad påverkar resultaten i svensk grundskola? Kunskapsöversikt om betydelsen av olika faktorer* (ss. 40–83). Stockholm: Skolverket.
- Harlen, W. & Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning. *Research Evidence in Educational Library*. Issue 1. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. London: Routledge.
- Hattie, J. A. C. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, 44(3), 513–524.
- Hobfoll, S. E. (2001). The influence of culture, community, and the nested-self in the stress process: Advancing conservation of resources theory. *Applied Psychology: An International Review*, 50(3), 337–421.
- Härnqvist, K. (2000). Evaluation through follow-up. A longitudinal program for studying education and career development. In C.-G. Janson (Ed.), *Seven Swedish Longitudinal Studies in Behavioural Science* (pp. 76–114). Stockholm: Forskningsrådsnämnden.
- Johnston, J. & McClune, W. (2000). *Selection project sel 5.1: Pupil motivation and attitudes – self-esteem, locus of control, learning disposition and the impact of selection on teaching and learning*. Belfast: Queen's University.
- Klapp Lekholm, A. & Cliffordson, C. (2009). Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation*, 15(1), 1–23.
- Klapp, A., Cliffordson, C. & Gustafsson, J.-E. (2014). The effect of being graded on later achievement – evidence from 13-Year Olds in Swedish Compulsory School. *Educational Psychology: An International Journal of Experimental Educational Psychology*, DOI: 10.1080/01443410.2014.933176.

- Levin, H. (2009). The economic pay off to investing in educational justice. *Educational Researcher*, 38(5), 5–20.
- Muthén, L. K. & Muthén, B. O. (2009). *Mplus User's Guide (5th edition)*. Los Angeles, CA: Muthén & Muthén.
- Muthén, B., Kaplan, D. & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52(3), 431–462.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22(2), 155–175.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144.
- Sjögren, A. (2010). *Graded children – evidence of long-run consequences of school grades from a nationwide reform*. Uppsala: IFAU.
- Svensson, A. (1971). Relative achievement. School performance in relation to intelligence, sex and home environment. Doctoral dissertation. *Göteborg studies in educational science, Vol. 6*. Göteborg: Acta Universitatis Gothoburgensis.
- Skolverket (2014). Sommarbetyg påverkar behörighet. Hämtad 2014-09-22 från <http://www.skolverket.se/press/pressmeddelanden/2014/sommarbetyg-paverkar-behorighet-1.222542>
- Skolverket (2014). Skolverkets slutsatser om likvärdighet och det fria skolvalet. Hämtad 2014-09-22 på <http://www.skolverket.se/statistik-och-utvardering/nyhetsarkiv/2.7602/skolverkets-slutsatser-om-likvardighet-och-det-fria-skolvalets-effekter-1.211468>
- Taylor, S. E. (1991). Asymmetric effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin*, 110(1), 67–85.
- Utbildningsdepartementet (2013). *Läroplaner med nya kursplaner och kunskapskrav samt en ny betygsskala*. Hämtad 2013-11-08 på <http://www.regeringen.se/content/1/c6/11/72/72/b821e577.pdf>
- Voyer, D. & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*. Advanced online publication. <http://dx.doi.org/10.1037/a0036620>