



**IFAU**

Institutet för arbetsmarknads- och  
utbildningspolitisk utvärdering

# **Tre lärdomar från en effektutvärdering av lärarstödsprogrammet NTA**

**Erik Mellander  
Joakim Svärdh**

**RAPPORT 2015:17**

Institutet för arbetsmarknads- och utbildningspolitisk utvärdering (IFAU) är ett forskningsinstitut under Arbetsmarknadsdepartementet med säte i Uppsala. IFAU ska främja, stödja och genomföra vetenskapliga utvärderingar. Uppdraget omfattar: effekter av arbetsmarknads- och utbildningspolitik, arbetsmarknadens funktionssätt och arbetsmarknadseffekter av socialförsäkringen. IFAU ska även sprida sina resultat så att de blir tillgängliga för olika intressenter i Sverige och utomlands.

IFAU delar även ut forskningsbidrag till projekt som rör forskning inom dess verksamhetsområden. Forskningsbidragen delas ut en gång per år och sista dag för ansökan är den 1 oktober. Eftersom forskarna vid IFAU till övervägande del är nationalekonomer, ser vi gärna att forskare från andra discipliner ansöker om forskningsbidrag.

IFAU leds av en generaldirektör. Vid institutet finns ett vetenskapligt råd bestående av en ordförande, institutets chef och fem andra ledamöter. Det vetenskapliga rådet har bl.a. som uppgift att lämna förslag till beslut vid beviljandet av forskningsbidrag. Till institutet är även en referensgrupp knuten där arbetsgivar- och arbetstagersidan samt berörda departement och myndigheter finns representerade.

Rapporterna finns även i tryckt format. Du kan beställa de tryckta rapporterna via telefon eller mejl. Se nedanstående kontaktinformation.

Postadress: Box 513, 751 20 Uppsala  
Besöksadress: Kyrkogårdsgatan 6, Uppsala  
Telefon: 018-471 70 70  
Fax: 018-471 70 71  
ifau@ifau.uu.se  
www.ifau.se

IFAU har som policy att en uppsats, innan den publiceras i rapportserien, ska seminariebehandlas vid IFAU och minst ett annat akademiskt forum samt granskas av en extern och en intern disputerad forskare. Uppsatsen behöver dock inte ha genomgått sedvanlig granskning inför publicering i vetenskaplig tidskrift. Syftet med rapportserien är att ge den ekonomiska politiken och den ekonomisk-politiska diskussionen ett kunskapsunderlag.

# Tre lärdomar från en effektutvärdering av lärarstödsprogrammet NTA<sup>a</sup>

av

Erik Mellander<sup>b</sup> och Joakim Svärth<sup>c</sup>

2015-10-19

## Sammanfattning

I en utvärdering av effekterna av programmet NTA (Naturvetenskap och teknik för alla) på elevers skolresultat i årskurs 9 finner vi positiva och statistiskt säkerställda effekter. Dessa effekter framträder dock endast när vi tar hänsyn till att deltagare i NTA och icke-deltagare systematiskt skiljer sig åt avseende t.ex. föräldrars utbildningsnivå. Effekterna är också begränsade till ett ämne, fysik, och hänför sig endast till resultaten på nationella prov – inga effekter kan spåras på betygen. Med hjälp av dessa resultat diskuterar vi tre metodologiska aspekter som är av generell intresse vid utvärderingar av interventioner på utbildningsområdet: att programdeltagare och icke-deltagare i allmänhet inte är jämförbara när det gäller initiala förutsättningar, att ett programs effekter kan skilja sig mellan kunskapsområden, samt betydelsen av standardiserade prestationsmått.

---

<sup>a</sup> Stort tack till Helena Holmlund för värdefulla kommentarer.

<sup>b</sup> Institutet för arbetsmarknads- och utbildningspolitisk utvärdering (IFAU), erik.mellander@ifau.uu.se

<sup>c</sup> KTH, Skolan för teknikvetenskaplig kommunikation och lärande, jsvardh@kth.se

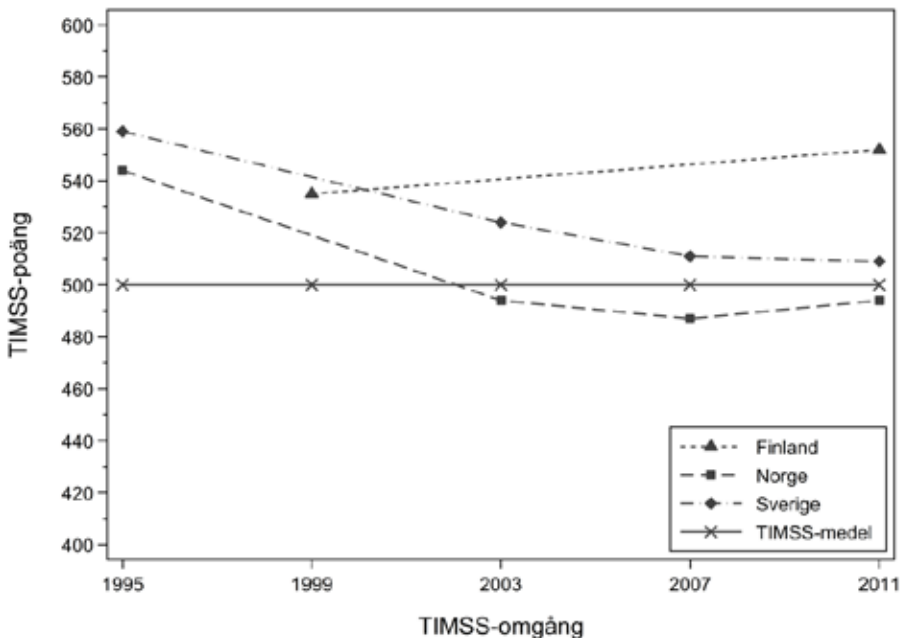
## Innehållsförteckning

|     |   |    |
|-----|---|----|
| 1   | Inledning.....  | 3  |
| 2   | Vad är NTA? .....   | 5  |
| 2.1 | Föregångaren: Science and Technology for Children (STC) .....     | 5  |
| 2.2 | Skillnader mellan NTA och STC.....                                | 6  |
| 2.3 | Anslutning till NTA.....  | 6  |
| 3   | Tidigare analyser av NTA .....                                    | 7  |
| 4   | Framtagandet av analysens dataunderlag .....                      | 7  |
| 4.1 | Stickprov från nationella prov i biologi, fysik, och kemi .....   | 8  |
| 4.2 | Uppdelningen av stickproven med avseende på NTA.....              | 9  |
| 4.3 | Tillförande av skol-, kommun- och länsdata .....                  | 10 |
| 5   | Resultat .....  | 10 |
| 5.1 | Betydelsen av att ha en jämförbar kontrollgrupp .....             | 11 |
| 5.2 | Skiljer sig effekten av NTA mellan biologi, fysik och kemi? ..... | 15 |
| 5.3 | Vikten av standardiserade utfallsmått.....                        | 16 |
| 6   | Sammanfattning och diskussion .....                               | 20 |
|     | Referenser .....  | 23 |
|     | Appendix.....   | 24 |

# 1 Inledning

Sveriges resultat i naturvetenskap på den internationella kunskapsundersökningen TIMSS (Trends in International Mathematics and Science Study) sjönk kraftigt mellan åren 1995 och 2003, se Figur 1. En sådan nedgång väcker diskussioner om behov av att öka intresset för naturvetenskap och höja kunskaperna i de naturvetenskapliga ämnena. Lärarstödsprogrammet NTA (Naturvetenskap och teknik för alla), som introducerades 1997, är det hittills mest ambitiösa och genomarbetade program som har utvecklats för detta ändamål.

Figur 1 Resultat i den naturvetenskapliga delen av TIMSS i årskurs 8: medelpoäng för Finland, Norge och Sverige



Noter:

1. Sverige och Norge deltog i TIMSS 1995, 2003, 2007 och 2011, Finland 1999 och 2011.

2. Alla länderskillnader är statistiskt säkerställda utom mellan Sverige och Norge 1995.

Källa: [timssandpirls.bc.edu](http://timssandpirls.bc.edu)

Av Figur 1 framgår att Sveriges resultat i TIMSS inte förbättrades efter 2003. Under perioden 2003–2011 föll de dock i långsammare takt än mellan 1995 och 2003, och sedan 2007 har fallet avstannat. Kan NTA ha något att göra med denna utveckling? Det kan vi inte veta, eftersom vi saknar information om vad

som hade hänt om NTA inte hade funnits – vi saknar information om det *kontrafaktiska* händelseförloppet.

Ett sätt att bilda sig en uppfattning om hur den kontrafaktiska utvecklingen hade kunnat se ut är att undersöka hur det har gått för länder som liknar Sverige, men som inte har något NTA-program. Resultaten för två sådana länder, Norge och Finland, framgår av Figur 1. I Norge har resultaten legat på en konstant nivå sedan 2003 – de små förändringarna i genomsnittresultaten är inte statistiskt säkerställda. Finland har endast deltagit i TIMSS två gånger, 1999 och 2011, och uppvisar mellan dessa två år en liten resultatförbättring.<sup>1</sup>

Både Norge och Finland har alltså haft en bättre utveckling än Sverige. En jämförelse med dessa två länder ger således inget stöd för att NTA skulle ha bidragit positivt till Sveriges resultat i TIMSS. Men det utesluter inte att NTA ändå kan ha haft en positiv effekt – det är inte säkert att Norge och Finland är bra jämförelseländer.

Ett annat sätt att undersöka NTA:s påverkan vore att genomföra analysen på individnivå för att kartlägga om elever som har deltagit i NTA klarar sig bättre i TIMSS än elever som inte har tagit del av NTA. Men detta skulle kräva att man bland eleverna som har testats i TIMSS kunde urskilja vilka som har deltagit i NTA, något som tyvärr är omöjligt. Även om den vore möjlig skulle en jämförelse mellan deltagare och icke-deltagare inte nödvändigtvis ge en rättvisande bild av programmets effekter: skolor och deras huvudmän väljer själva om de vill arbeta med naturvetenskap och teknik inom ramen för NTA, vilket kan innebära att det finns systematiska skillnader mellan NTA-elever och elever som inte har deltagit i NTA. Eleverna kan t.ex. skilja sig åt med avseende på föräldrarnas utbildningsbakgrund, vilket är en faktor som i sig är starkt korrelerad med elevens studieresultat. Dessa systematiska skillnader måste man ta hänsyn för att jämförelsen ska bli rättvisande.

Det är dock möjligt att lösa dessa problem, med andra data än TIMSS. Vi har nyligen genomfört den första effektutvärderingen av NTA: en statistisk analys där vi undersöker orsakssambandet mellan resultat i årskurs 9 på nationella prov i de naturvetenskapliga ämnena 2009–2010 och tidigare deltagande i NTA, se Mellander och Svärth (2015). Genom att ta hänsyn till faktorer som bestämmer vilka elever som får ta del av NTA kan vi jämföra NTA-deltagare med elever som inte har tagit del av NTA *som om* deltagandet i NTA hade bestämts slumpmässigt, i analogi med ett medicinskt experiment. På så

---

<sup>1</sup> De två övriga nordiska länderna, Danmark och Island, har endast deltagit en gång i den naturvetenskapliga delen av TIMSS för årskurs 8, år 1995, varför någon utveckling över tiden inte kan redovisas för dessa länder.

sätt säkerställs att de två grupperna är jämförbara. Detta visar sig ha stor betydelse för våra resultat.

Vi finner också att vi inte kan uttala oss om effekterna av NTA på färdigheter i naturvetenskap i allmänhet. Våra slutsatser skiljer sig mellan olika naturvetenskapliga ämnen. Dessutom är de beroende av att vi mäter effekterna i termer av resultat på de nationella proven – resultaten blir annorlunda när vi mäter effekterna av NTA på kursbetygen i årskurs 9.

Med vår effektutvärdering som utgångspunkt diskuterar – och illustrerar – vi i denna artikel tre metodologiska aspekter som är av allmänt intresse vid utvärderingar av interventioner på utbildningsområdet: att programdeltagare och icke-deltagare i allmänhet inte är jämförbara, att ett programs effekter kan skilja sig mellan kunskapsområden, samt betydelsen av standardiserade prestationsmått. Artikeln är disponerad som följer.

Avsnitt 2 innehåller en beskrivning av NTA och hur det går till att ansluta sig till programmet. Därefter, i avsnitt 3, sammanfattar vi tidigare analyser av NTA. I avsnitt 4 redogör vi för hur vi har sammanställt det datamaterial som vi använder. Resultaten av vår analys diskuteras i avsnitt 5 och avsnitt 6 innehåller en sammanfattning och en avslutande diskussion.

## 2 Vad är NTA?

NTA introducerades år 1997 och är den svenska versionen av det amerikanska lärarstödsprogrammet Science and Technology for Children (STC). Vid en beskrivning av NTA är det därför naturligt att utgå från en beskrivning av STC.

### 2.1 Föregångaren: Science and Technology for Children (STC)

Tidiga versioner av STC kan spåras ända tillbaka till 1960-talet. I sin nuvarande form lanserades programmet 1995. Programmet innehåller tre komponenter: läromaterial – ”experimentlådor” – med tillhörande instruktioner, lärarhandledning och -utbildning, samt stöd till skollära.

STC är uppdelat på två block. Det första är avsett för elever från förskolan och upp till årskurs 6. Det andra blocket är avsett för årskurs 6–8. Inom blocken är programmet dessutom ålders- och ämnesindelad. Ämnesområdena är: Life Science<sup>2</sup>, Earth Science<sup>3</sup> samt Physical Science and Technology.

---

<sup>2</sup> Life Science är en tvärvetenskaplig forskningsgren, som innefattar bland annat innefattar biologi, medicin, fysiologi och zoologi.

<sup>3</sup> Earth Science är också en tvärvetenskaplig forskningsgren. Den kombinerar ämnena fysik, kemi, biologi och matematik.

Det finns ett drygt 30-tal olika teman att välja mellan i STC. Temana är tänkta att sträcka sig över en termin och föregås av en heldagsutbildning för den undervisande läraren. Arbetet med temat följer en mall där eleverna under lärarens ledning får formulera hypoteser, utföra experiment, analysera resultaten och skriftligt dokumentera sitt arbete och sina slutsatser.

## 2.2 Skillnader mellan NTA och STC

NTA är på svenska och introducerades 1997. Det var Kungliga Vetenskapsakademien (KVA) som i samarbete med Kungliga Ingenjörsvetenskapsakademien (IVA) översatte STC-programmet och anpassade det till den svenska läroplanen, Lpo94.<sup>4</sup> NTA är i första hand avsett för årskurs 1–6 men har också använts i årskurs 7–9.

Antalet teman, och därmed antalet experimentlådor, är något lägre i NTA än i STC – under den period vi studerar uppgick antalet NTA-teman till drygt 20 stycken. NTA innehåller inte heller lika utförligt stöd för skolledare som STC. Därutöver är den lärarutbildning som är kopplad till NTA något mindre omfattande än motsvarande lärarutbildning inom ramen för STC. Lärarnas arbete med NTA stöds av lokala NTA-samordnare som också utgör länkar till regionala NTA-organisationer och till NTA Skolutveckling som skapar och utvecklar det rikstäckande NTA-programmet.

## 2.3 Anslutning till NTA

Både kommunala och privata skolor kan använda sig av NTA. För att skolorna ska kunna göra det måste skolornas huvudmän dock vara medlemmar i NTA Skolutveckling.<sup>5</sup> I samband med att skolhuvudmännen ansluter sig erlägger de en inträdesavgift på 15 000 kronor.

Skolhuvudmannen beslutar, ofta i samråd med de skolor som omfattas av huvudmannskapet, vilka skolor, eller delar av skolor, som ska ansluta sig till NTA.

Skolor som deltar i NTA betalar en fast kostnad per experimentlåda som de använder, cirka 5 000 kronor. Dessutom betalar de en terminsavgift på 20 kronor per deltagande elev.

År 2010 användes NTA i ungefär en tredjedel av Sveriges 290 kommuner. Antalet deltagande lärare och elever uppgick till cirka 6 700 respektive 96 000. Eftersom grundskolan omfattade ungefär 1 miljon elever innebär detta att 10

---

<sup>4</sup> Den nuvarande versionen är anpassad till Lgr11.

<sup>5</sup> För kommunala skolor är det naturligtvis kommunen som är huvudman. För privata skolor kan skolan själv vara huvudman, om skolan är enskild, eller ett bolag, om skolan ingår i en koncern.



procent av grundskoleeleverna deltog i NTA. Vid den tiden uppgick den årliga tillväxttakten i verksamheten till omkring 10 kommuner och 10 000 elever.

### **3 Tidigare analyser av NTA**

Det saknas tidigare effektutvärderingar av NTA, d.v.s. analyser syftande till att fastställa orsakssamband mellan deltagande i NTA och färdigheter i naturvetenskap och teknik. Det finns inte heller några tidigare kvantitativa uppföljningsstudier.

Däremot har ett flertal kvalitativa analyser av NTA genomförts. Dessa visar genomgående på positiva erfarenheter av programmet. Exempelvis finner Gisselberg (2001), Schoultz och Hultman (2002) samt Schoultz, Hultman och Lindkvist (2003) att lärarna i huvudsak är nöjda med den information de har fått om NTA och den undervisning och träning de har erhållit, samt att eleverna är entusiastiska till NTA.

Anderhag och Wickman (2006) fann att lärare som använde NTA tyckte att de därigenom förbättrade sin förmåga att stödja elevernas språkveckling och förståelse av naturvetenskapliga begrepp. Denna studie följdes upp av en intervjustudie omfattande 80 elever i årskurs 6, se Anderhag och Wickman (2007). Av de 80 eleverna valdes 40 stycken slumpmässigt bland NTA-deltagare och 40 stycken slumpmässigt bland elever som inte deltagit i NTA. Slutsatsen var att NTA-deltagarna hade fått en djupare förståelse av naturvetenskapliga begrepp och processer än de elever som inte hade deltagit i NTA.

Studien av Gisselberg (op.cit.) innehåller en intressant observation med avseende på de olika ämnesområden som NTA omfattar. Merparten av de lärare som använde NTA tyckte sig ha tillräckliga sakkunskaper för att tillämpa undervisningsmaterialet i biologi men kände sig mindre säkra på sin kompetens inom fysik, kemi och teknologi.

### **4 Framtagandet av analysens dataunderlag**

Det datamaterial vi har använt har byggts upp i två steg. Det första steget utgörs av två stickprov från resultaten i nationella prov i naturvetenskapliga ämnen 2009 och 2010. I det andra steget klassificeras individerna i dessa stickprov med avseende på om de deltagit i NTA eller inte. I det sista steget påförs skol-, kommun- och länsinformation. De tre stegen beskrivs i tur och ordning.

#### 4.1 Stickprov från nationella prov i biologi, fysik, och kemi

Nationella prov i årskurs 9 i biologi, fysik och kemi introducerades år 2009. Alla elever i årskurs 9 omfattas av proven men en elev i en given skola gör endast provet i ett av de tre ämnena. Ämnesproven fördelas på skolor av Skolverket enligt en modell som i princip är slumpmässig, men som ser till att inga skolor får prov i samma ämne två år i rad.

Forskare vid Institutionen för tillämpad utbildningsvetenskap vid Umeå universitet utformade proven. För att följa upp resultaten beslöt man sig också för att dra ett slumpmässigt stickprov av resultaten bland de elever som skrev provet. Stickprovet omfattade alla elever födda på tre specifika dagar i månaden, vilket innebar att stickprovet skulle motsvara 10 procent av eleverna i årskurs 9. Förutom resultaten på de nationella proven – provpoäng och provbetyg – samlade man också in uppgifter om elevernas kursbetyg, deras kön samt om de deltagit i hemspråksundervisning.

I praktiken kom Umeå universitets stickprov att omfatta en (betydligt) lägre andel än 10 procent av eleverna. Detta framgår av Tabell 1.

Tabell 1 Totala antalet elever i årskurs 9 läsåren 2008/2009 och 2009/2010, antal som skrev nationella prov i biologi, fysik eller kemi, samt antal som ingick i Umeå universitets stickprov

|              | Antal elever<br>i årskurs 9 | Antal (%) som skrev<br>nationellt prov i bi/fy/ke | Varav antal (%) ingick i Umeå<br>universitets stickprov |
|--------------|-----------------------------|---|---|
| 2008/2009    | 118 032                     | 88 491 (75,0)                                     | 8 028 (9,1)   |
| 2009/2010    | 113 545                     | 98 848 (87,1)                                     | 7 839 (7,9)   |
| <i>Summa</i> | <i>231 577</i>              | <i>187 339 (80,9)</i>                             | <i>15 867 (8,5)</i>                                     |

Källor: Skolverket; Inst. för tillämpad utbildningsvetenskap, Umeå universitet; egna beräkningar.

Den andra kolumnen i Tabell 1 visar att betydande andelar av eleverna inte skrev det nationella provet i de naturvetenskapliga ämnena – 25 procent 2008/2009 och knappt 13 procent 2009/2010. Detta berodde på att vissa skolor avstod från att genomföra provet, särskilt under läsåret 2008/2009, trots att deltagande var obligatoriskt. Den tredje kolumnen visar att bland dem som skrev provet var det en lägre andel än 10 procent som kom att ingå i Umeå universitets stickprov. Förklaringen är att vissa skolor inte rapporterade in sina resultat till Umeå universitet – motsvarande 0,9 procent av de elever som skrev provet 2008/2009 och 2,1 procent 2009/2010. Att detta kunde ske berodde på att universitet saknade möjligheter att tvinga de deltagande skolorna att rapportera sina resultat. Sammantaget ledde bortfallet i form av skolor som inte skrev provet respektive skolor som skrev provet men inte rapporterade resultaten till

att stickproven kom att uppgå till knappt 7 procent av eleverna i årskurs 9 båda läsåren (8 028/118 032 respektive 7 839/113 545).

Hur kan det bortfall som framgår av Tabell 1 påverka våra resultat? Det är rimligt att tro att resultaten i bortfallsgruppen i genomsnitt är lägre än resultaten bland dem som ingår i stickproven – skolor där eleverna har presterat sämre bör vara mindre benägna att rapportera in sina resultat än skolor med goda resultat. Om NTA har haft en positiv effekt på elevernas kunskaper bör bortfallet i så fall vara lägre bland skolor vars elever deltagit i NTA, än i övriga skolor. Detta kommer att medföra en underskattning av effekterna av NTA. Slutsatsen blir att om våra skattningar visar på positiva effekter av NTA kan de ses som en undre gräns för de sanna effekterna.

## 4.2 Uppdelningen av stickproven med avseende på NTA

Utgångspunkten för klassificeringen av enskilda elever med avseende på NTA är uppgifter på skolnivå. Med hjälp av NTA-organisationen har samtliga skolor som hade elever i årskurs 9 år 2009 – drygt 1 850 skolor – hänförts till en av följande tre kategorier:

- i. Skolans samtliga elever hade deltagit i NTA. Dessa skolor utgjorde knappt 100 stycken. I merparten av dem, 73 stycken, hade genomsnittseleven deltagit i NTA under 3–4 terminer upp till och med årskurs 6. I vissa skolor, 23 stycken, hade eleverna också deltagit i NTA under någon eller några av årskurserna 7–9.
- ii. Inte någon av skolans elever hade deltagit i NTA. Dessa skolor uppgick till 1 555 stycken.
- iii. Vissa, men inte alla, av skolans elever hade deltagit i NTA *eller* information saknades om huruvida skolans elever hade deltagit i NTA. Sammanlagt rörde det sig om knappt 200 stycken.

I vår analys bygger klassificeringen av elever direkt på denna klassificering av skolor. Det innebär att elever vars skolor tillhör kategori i. klassificeras som NTA-elever och att elever vars skolor tillhör kategori ii. klassificeras som icke-NTA elever, medan elever vars skolor tillhör kategori iii. inte kan klassificeras varken som NTA- eller icke-NTA-elever.

Genom att matcha denna klassificering med eleverna i den sista kolumnen i Tabell 1 erhåller vi följande uppdelning av eleverna:

Tabell 2 Uppdelningen av Umeå universitets stickprov efter NTA-status

| NTA-elever | Icke-NTA elever | Studerade elever:<br>NTA + icke-NTA | Ej klassificerade elever | Summa  |
|------------|-----------------|-------------------------------------|--------------------------|--------|
| 1 121      | 11 845          | 12 966                              | 2 901                    | 15 867 |

Av de 15 867 elever som ingick i Umeå universitets stickprov används således 12 966 stycken i analyserna nedan – 2 901 elever faller bort eftersom de inte kan klassificeras som antingen NTA- eller icke-NTA-elever.

Det ska noteras att även bland de 12 966 elever som ingår i våra analyser föreligger det viss osäkerhet om de har deltagit i NTA eller inte. Denna osäkerhet beror på att elever kan ha bytt skola. På en skola vars elever har klassificerats som NTA-elever kan det finnas elever som inte har deltagit i NTA, men som har hamnat där efter byte av skola. På motsvarande sätt kan det på skolor vars elever har klassificerats som icke-NTA-elever finnas elever som har deltagit i NTA.<sup>6</sup> Det finns dock ingen anledning att tro att NTA är en viktig faktor vid skolbyten. Felklassificeringarna kan därför betraktas som slumpmässiga mätfel. Dessa medför att *om* den sanna effekten av NTA är positiv kommer den att underskattas, d.v.s. påverkas på samma sätt som av bortfallet i stickprovsundersökningarna, jfr. avsnitt 4.1.

### 4.3 Tillförande av skol-, kommun- och länsdata

Information på skolnivå har inhämtats ifrån Skolverkets databas SALSA (Skolverkets Arbetsverktyg för Lokala SambandsAnalyser). Uppgifterna avser dels skolans elever – andel pojkar, andel utlandsfödda, andel svenskfödda med utlandsfödda föräldrar, andel med minst godkänt betyg i alla ämnen i årskurs 9, dels den genomsnittliga utbildningsnivån hos elevernas föräldrar.

Data på kommunnivå har erhållits från Sveriges kommuner och landstings databas Kolada (Kommun- och landstingsdatabasen). Informationen omfattar demografi – befolkningsstorlek och befolkningstäthet, ekonomi – medianinkomst och utbildning – lärartäthet, andel behöriga lärare, kostnader per elev i årskurserna 1–9 samt andel elever i årskurs 9 med minst godkänt betyg i alla ämnen.

Länsdata, slutligen, utgörs av uppgifter om i vilka län NTA- och icke-NTA-elevernas skolor är belägna.

## 5 Resultat

Detta avsnitt har två syften, ett specifikt och ett generellt. Det specifika syftet är att redovisa resultaten av vår utvärdering av NTA:s effekter på skolresultat i årskurs 9. Det generella syftet är att med hjälp av dessa resultat illustrera tre allmänna aspekter på utvärderingar av effekter av utbildningssatsningar: skillnaden mellan en bra och en dålig jämförelsegrupp, skillnaden mellan genom-

---

<sup>6</sup> Skolbyten *behöver* dock inte innebära att elever blir felaktigt klassificerade – elever kan ju ha bytt mellan NTA-skolor eller mellan icke-NTA-skolor.

snittliga effekter och effekter inom olika kunskapsområden, samt skillnaden mellan standardiserade och icke standardiserade mått på skolprestationer. Avsnittet är strukturerat utifrån det generella syftet.

## 5.1 Betydelsen av att ha en jämförbar kontrollgrupp

Ett enkelt sätt att undersöka om deltagande i NTA verkar ha haft någon långsiktig effekt på elevernas skolprestationer är att undersöka om de elever som deltagit i NTA i genomsnitt presterar bättre än de elever som inte har deltagit i NTA. För att testa detta utnyttjar vi det mest precisa måttet på skolprestationer som vi har tillgång till, nämligen poäng på de nationella proven i biologi, fysik och kemi.

Vi kan inte använda provpoängen direkt som den är eftersom det maximala antalet poäng skiljer sig mellan de olika ämnena och även mellan åren. För att göra individernas provpoäng jämförbara mellan ämnen och över år har vi räknat om provpoängen till *percentilrangordningar*. Det är ett vanligt tillvägagångssätt som innebär att vi ersätter elevens provpoäng med hans placering i fördelningen av resultaten för det aktuella provet, där fördelningen går mellan 0 och 100. Den individ som hade det högsta resultatet på provet får percentilrangen 100, den som har precis lika många före sig som efter sig får percentilrangen 50, den som har det lägsta resultat får percentilrangen 0 o.s.v.<sup>7</sup>

I Tabell 3 jämför vi de genomsnittliga percentilrangordnade provpoängen bland dem som har deltagit i NTA med dem som inte har deltagit. Denna jämförelse tyder inte på några positiva effekter av NTA. Tvärtom, tabellen visar att den genomsnittliga percentilrangen bland dem som har deltagit i NTA är lägre än bland icke-deltagarna och att denna skillnad är statistiskt säkerställd.

Tabell 3 Percentilrangordnade (p-rangordnade) resultat på nationella prov i biologi, fysik och kemi: NTA-deltagare och icke-NTA-deltagare

| NTA-deltagare   |              |                    | Icke-NTA-deltagare |              |                    | NTA – icke-NTA         |                    |
|-----------------|--------------|--------------------|--------------------|--------------|--------------------|------------------------|--------------------|
| Antal individer | Medel p-rang | Standard-avvikelse | Antal individer    | Medel p-rang | Standard-avvikelse | Differens <sup>1</sup> | Standard-avvikelse |
| 1 121           | 48,71        | 27,86              | 11 845             | 50,56        | 28,88              | -1,85*                 | 0,900              |

<sup>1</sup> Differens = medelvärdet för NTA-deltagare – medelvärdet för individer som ej deltagit i NTA. Symbolerna \* och \*\* markerar att ett dubbelsidigt *t*-test visar att differensen är statistiskt säkerställd på signifikansnivåerna 5 respektive 1 procent.

<sup>7</sup> Percentilrangen beräknas enligt  $[(c_t + 0,5f_t) / N] \times 100$ , där  $c_t$  är antalet resultat som är lägre än det aktuella, *i*.te, resultatet,  $f_t$  är antalet personer som har det aktuella resultatet och  $N$  är antalet personer som gjorde provet.

Vad kan ligga bakom resultatet i Tabell 3? Den rimligaste förklaringen har att göra med vad som avgör vilka individer som deltar i NTA och vilka som inte gör det. Testet som redovisas i Tabell 3 är berättigat bara under ett viktigt underförstått antagande, nämligen att NTA-deltagarna och icke-deltagarna är jämförbara så tillvida att de i genomsnitt är lika i alla avseenden som är relevanta för deras studieresultat, och att det enda som skiljer dem åt är att de har, respektive inte har, deltagit i NTA. Detta antagande innebär t.ex. att det inte föreligger några systematiska skillnader när det gäller elevernas socioekonomiska bakgrund, migrationsbakgrund, tidigare skolresultat och studiemotivation.

Eftersom skolor själva väljer att delta i NTA är det rimligt att det uppstår systematiska skillnader mellan NTA-deltagarna och icke-NTA-individerna. Om dessa skillnader inte beaktas vid utvärderingen kommer de skattade effekterna av NTA att bli felaktiga.<sup>8</sup>

Vilka skillnader mellan dem som deltagit i NTA och de som inte har gjort det skulle kunna ge upphov till de till synes negativa effekter av NTA som vi just har noterat? En möjlig förklaring är att NTA har använts för att stärka undervisningen i skolor där eleverna presterar sämre än genomsnittet. Det finns vissa tecken i våra data på att det skulle kunna vara så: andelen elever med godkänt i alla ämnen är lägre i NTA-skolor än i andra skolor och i NTA-skolor har elevernas föräldrar lägre utbildning än föräldrarna till barn som går i skolor som inte använder NTA, se Mellander och Svärth (2015).<sup>9</sup> I så fall kommer NTA-deltagande att vara överrepresenterat bland skolor med låga resultat på nationella prov och låga betyg. Detta måste beaktas när NTA:s effekter ska fastställas, annars kommer effekterna att underskattas.

För att kunna göra en rättvis jämförelse måste vi i våra data välja ut NTA-deltagare och icke-NTA-individer som är så lika som möjligt i alla andra avseenden än just ifråga om NTA. Detta är inte alldeles enkelt eftersom NTA- och icke-NTA-individer kan skilja sig åt på många olika sätt. Det räcker inte med att hitta individer som, exempelvis, har samma kön och föräldrar med samma utbildningsnivå.

En metod som samtidigt kan beakta många olika skillnader mellan NTA-deltagarna och de individer som inte har deltagit i NTA är ”propensity score matching”, PSM, se Guo och Fraser (2010). Denna metod innebär att man med statistiska metoder skattar sannolikheten – propensity score – att elever har del-

---

<sup>8</sup> I utvärderingslitteraturen kallas detta systematiska fel i effektskattningen för ”selektionsbias”, jfr. Guo och Fraser (2010, s. 2324), vilket syftar på att felet beror på att man inte tar hänsyn till (den icke slumpmässiga) process som styr urvalet (selektionen) till ”behandling”.

<sup>9</sup> Men det finns också data som pekar i motsatt riktning. Exempelvis är andelen utlandsfödda elever lägre i NTA-skolor än i andra skolor, se Mellander och Svärth (op.cit.).

tagit i NTA, som en funktion av flera olika variabler. Dessa skattningar kan göras *både för elever som faktiskt har gått NTA och elever som inte har gjort det.*<sup>10</sup> Man kan visa att elever med samma propensity score, d.v.s. samma sannolikhet att delta i programmet, kan betraktas som lika i alla de dimensioner som variablerna representerar.<sup>11</sup>

Idén med PSM är att skapa en kontrollgrupp genom att till varje NTA-deltagare hitta icke-deltagare som har samma, eller nästan samma, propensity score. Jämförelser mellan sådana par av deltagare och icke-deltagare kan tolkas ”som om” fördelningen på behandlingsgrupp (NTA) och kontrollgrupp (icke-NTA) har varit slumpmässig och därmed gett upphov till jämförbara grupper. Här finns dock ett viktigt förbehåll: propensity score-skattningen måste inkludera samtliga faktorer som har betydelse för om en individ tar del av behandlingen. Detta krav är i allmänhet svårt att uppfylla eftersom det ofta kan finnas viktiga faktorer som är svåra att observera i data, som t.ex. attityder och ambitioner. I sådana situationer kan det hända att resultaten från PSM inte är tillförlitliga, och att de bör tolkas med försiktighet. För att undgå problemet med att programdeltagare och kontrollgrupp inte är jämförbara i alla avseenden kan effektutvärderingar också genomföras med hjälp av andra metoder, som dock kräver att den insats som ska utvärderas utformas på ett sätt som underlättar utvärdering. Vi återkommer till en diskussion om dessa i avsnitt 6.

Av avsnitt 2.3 framgår att valet att delta i NTA sker på kommun- och skolnivå. Detta innebär att icke observerbara individfaktorer utgör ett betydligt mindre problem för utvärderingen än om valet att delta hade fattats av eleverna själva. Men det eliminerar inte problemet helt – kommun- och skollidningens beslut kan påverkas av egenskaper hos eleverna som är kända för dem men som inte kan observeras av oss forskare. Resultaten i det följande bygger därför på antagandet att NTA-elever och icke-NTA elever som är jämförbara i termer av propensity scores också är jämförbara i termer av icke-observerade faktorer som t.ex. motivation och studieförmåga.

Resultatet av den propensity score-regression som vi har skattat redovisas i Tabell A1 i Appendix. Av tabellen framgår det att alla de variabler på skol- och

---

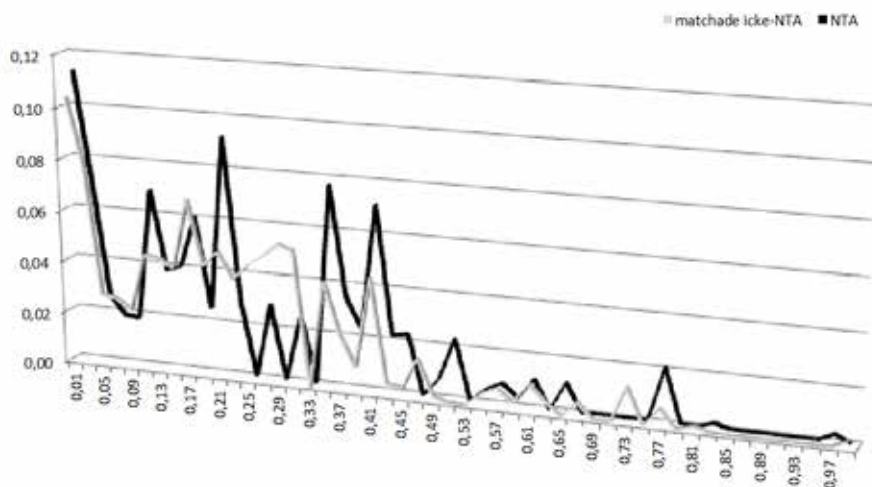
<sup>10</sup> Mera specifikt används så kallad logistiska regressioner där den beroende variabeln sätts till 1 för elever som har deltagit i NTA och till 0 för elever som inte har deltagit. De predikterade värden på den beroende variabeln som regressionen resulterar i utgörs av (reella) tal i intervallet (0,1), vilka kan tolkas som sannolikheten för att den aktuella individen ska ha deltagit i NTA, givet de förklaringsvariabler som ingår i regressionen.

<sup>11</sup> Detta innebär inte att en NTA-deltagare och en icke-deltagare med samma propensity score har likadana värden på alla variabler som ligger till grund för deras gemensamma propensity score men det innebär att om skillnader föreligger så är de inte systematiska utan slumpmässiga.

kommunnivå som har inkluderats i regressionen har ett statistiskt säkerställt samband med huruvida en elev har gått i en NTA-skola eller inte.

På basis av propensity score-skattningen har vi valt ut 1 000 par av NTA- och matchande icke-NTA-individer. Individerna i dessa par är jämförbara i den meningen att deras propensity scores ligger nära varandra. Figur 2 visar de relativa frekvenserna av olika propensity scores för de 1 000 NTA-individerna och de 1 000 icke-NTA-individerna. Av figuren framgår att de relativa frekvenserna av olika propensity scores är mycket lika i NTA-gruppen och den matchande gruppen av icke-NTA-individer. Bortsett från enstaka avvikelser, t.ex. i intervallet 0,25–0,29, ligger de två kurvorna påfallande nära varandra. Som framgår av noten till figuren är korrelationen mellan de två gruppernas propensity scores 0,991.

Figur 2 Relativa frekvenser av propensity scores – skattade sannolikheter för deltagande i NTA för faktiska deltagare och matchade icke-deltagare



Not: Korrelationen mellan propensity scores för NTA och icke-NTA är 0,991.

När vi undersöker skillnaderna i resultat mellan NTA- och icke-NTA-individer bland de 1 000 jämförbara paren blir slutsatsen en helt annan än i Tabell 3. Det resultat vi får nu redovisas i Tabell 4, av vilken det framgår att skillnaden i genomsnittlig percentilrang mellan NTA-deltagare och icke-NTA-deltagare är signifikant positiv, på signifikansnivån 1 procent.



Tabell 4 Percentilrangordnade (p-rangordnade) resultat på nationella prov i biologi, fysik och kemi för NTA-deltagare och icke-NTA- deltagare som parats ihop (matchats) med hjälp av skattade propensity scores

| NTA-deltagare   |              |                    | Icke-NTA-deltagare |              |                    | NTA – icke-NTA         |                    |
|-----------------|--------------|--------------------|--------------------|--------------|--------------------|------------------------|--------------------|
| Antal individer | Medel p-rang | Standard-avvikelse | Antal individer    | Medel p-rang | Standard-avvikelse | Differens <sup>1</sup> | Standard-avvikelse |
| 1 000           | 48,42        | 28,07              | 1 000              | 44,38        | 28,44              | 4,04**                 | 1,263              |

<sup>1</sup> Differens = medelvärdet för NTA-deltagare – medelvärdet för individer som ej deltagit i NTA. Symbolerna \* och \*\* markerar att ett dubbelsidigt *t*-test visar att differensen är statistiskt säkerställd på signifikansnivåerna 5 respektive 1 procent.

De fullständigt olika slutsatser som Tabell 3 och Tabell 4 medför visar tydligt på betydelsen av att ha en jämförbar kontrollgrupp när man ska utvärdera effekter av utbildningsinsatser.

Resultatet i Tabell 4 innebär att jämfört med i övrigt lika elever som inte har deltagit i NTA så ligger NTA-deltagarna i genomsnitt fyra percentiler högre upp i fördelningen av resultaten på de nationella proven i naturvetenskapliga ämnen i årskurs 9. Om än positiv, kan denna skillnad förefalla ganska beskedlig. Vi ska nu titta närmare på den.

## 5.2 Skiljer sig effekten av NTA mellan biologi, fysik och kemi?

Den positiva NTA-effekt som redovisas i Tabell 4 är en aggregerad effekt som inte gör någon skillnad mellan de tre olika naturvetenskapliga ämnena. Det kan dock vara så att NTA har olika effekter på biologi, fysik och kemi. För att undersöka detta bryter vi ner NTA-effekten på de olika ämnena, jfr. Tabell 5.

Av Tabell 5 framgår att den positiva NTA-effekten hänför sig till fysikämnet. Skillnaderna i genomsnittlig percentilrang mellan NTA- och icke-NTA-deltagare är förvisso positiva även för biologi och kemi men för dessa ämnen är skillnaderna inte statistiskt säkerställda. Tabellen visar också att den aggregerade effekten för samtliga ämnen, naturligt nog, är betydligt mindre än effekten i fysik. Effekten i fysik, knappt sju percentiler, är betydande. Den innebär exempelvis att om en elev som inte har gått NTA har bättre resultat än en 1/4 av de övriga eleverna (25:e percentilen) så har den jämförbara NTA-eleven bättre resultat än knappt en 1/3 av de andra eleverna (32:a percentilen).

Tabell 5 Percentilrangordnade (p-rangordnade) resultat på nationella prov i biologi, fysik och kemi för NTA-deltagare och icke-NTA-deltagare som parats ihop (matchats) med hjälp av skattade propensity scores

| Ämne            | NTA-deltagare   |              |                    | Icke-NTA-deltagare |              |                    | NTA – icke-NTA         |                    |
|-----------------|-----------------|--------------|--------------------|--------------------|--------------|--------------------|------------------------|--------------------|
|                 | Antal individer | Medel p-rang | Standard-avvikelse | Antal individer    | Medel p-rang | Standard-avvikelse | Differens <sup>1</sup> | Standard-avvikelse |
| <i>Biologi</i>  | 339             | 45,38        | 28,40              | 339                | 43,88        | 26,85              | 1,50                   | 2,120              |
| <i>Fysik</i>    | 322             | 49,75        | 27,45              | 313                | 42,47        | 29,14              | 7,28**                 | 2,242              |
| <i>Kemi</i>     | 339             | 50,18        | 28,15              | 348                | 46,60        | 29,23              | 3,58                   | 2,187              |
| <i>Samtliga</i> | 1 000           | 48,42        | 28,07              | 1 000              | 44,38        | 28,44              | 4,04**                 | 1,263              |

<sup>1</sup> Differens = medelvärdet för NTA-deltagare – medelvärdet för individer som ej deltagit i NTA. Symbolerna \* och \*\* markerar att ett dubbelsidigt *t*-test visar att differensen är statistiskt säkerställd på signifikansnivåerna 5 respektive 1 procent.

I avsnitt 3 framkom ett resultat som är konsistent med Tabell 5: Gisselberg (2001) noterade att bland lärare som använde NTA tyckte sig de flesta ha bättre förkunskaper i biologi än i fysik och kemi. Följaktligen bör NTA ha haft större potential att förbättra undervisningen i fysik och kemi än i biologi. Denna hypotes stöds av Tabell 5 genom att den näst sista kolumnen visar på större positiva differenser för fysik och kemi än för biologi. För att närmare kunna diskutera vad skillnaderna i NTA-effekter mellan de olika naturvetenskapliga ämnena kan bero på behöver vi dock tillgång till mer information än vad som varit tillgänglig för denna studie. Vi lämnar därför denna analys som en uppgift för framtida forskning.

### 5.3 Vikten av standardiserade utfallsmått

I den internationella vetenskapliga litteraturen framhålls fördelarna med standardiserade utfallsmått när utbildningsinsatser ska utvärderas. Argumentet är att standardiserade utfallsmått innebär att skolprestationerna blir mer jämförbara än om de mäts i termer av resultat som (delvis) kan vara olika definierade och operationaliserade i olika skolor, se t.ex. Geier m.fl. (2008). I denna studie har vi dragit nytta av att införandet av nationella prov i de naturvetenskapliga ämnena år 2009 innebar att standardiserade utfallsmått för dessa ämnen blev tillgängliga, för första gången – de nationella proven är lika för alla i årskurs 9 och ska rättas efter enhetliga instruktioner.

Men hade det spelat någon roll om vi istället hade utvärderat effekterna av NTA genom deras inverkan på eleverna kursbetyg i årskurs 9? Detta är den fråga vi besvarar i detta avsnitt.

Av Tabell 6 framgår att vi inte hittar några statistiskt säkerställda effekter av NTA på kursbetyg, varken för de tre naturvetenskapliga ämnena sammantagna

eller var för sig. Det spelar alltså stor roll om man utvärderar NTA-effekten i termer av resultat på nationella prov eller i termer av kursbetyg.

Tabell 6 Kursbetyg i årskurs 9 i biologi, fysik och kemi för NTA-deltagare och icke-NTA-deltagare som parats ihop (matchats) med hjälp av skattade propensity scores

| Ämne            | NTA-deltagare    |                         |                   | Icke-NTA-deltagare |                         |                   | NTA – icke-NTA         |                   |
|-----------------|------------------|-------------------------|-------------------|--------------------|-------------------------|-------------------|------------------------|-------------------|
|                 | Antal individer  | Medelbetyg <sup>1</sup> | Standardavvikelse | Antal individer    | Medelbetyg <sup>1</sup> | Standardavvikelse | Differens <sup>3</sup> | Standardavvikelse |
| <i>Biologi</i>  | 339              | 12,17                   | 5,053             | 335                | 11,87                   | 4,646             | 0,30                   | 0,373             |
| <i>Fysik</i>    | 321              | 11,95                   | 4,351             | 312                | 11,94                   | 5,120             | 0,01                   | 0,377             |
| <i>Kemi</i>     | 335              | 12,04                   | 4,886             | 343                | 11,94                   | 4,937             | 0,10                   | 0,377             |
| <i>Samtliga</i> | 995 <sup>2</sup> | 12,06                   | 4,776             | 990 <sup>2</sup>   | 11,91                   | 4,895             | 0,15                   | 0,217             |

<sup>1</sup> Vid beräkningen av betygsmedelvärden har betygen värderats enligt: IG = 0, G = 10, VG = 15, MVG = 20.

<sup>2</sup> Antalet elever med kursbetyg är något lägre än de 1 000 elever vardera som utgör de matchade NTA- och icke-NTA-deltagarna, p.g.a. att några elever inte uppfyllde kraven för att kunna få kursbetyg.

<sup>3</sup> Differens = medelvärdet för NTA-deltagare – medelvärdet för individer som ej deltagit i NTA. Symbolerna \* och \*\* markerar att ett dubbelsidigt t-test visar att differensen är statistiskt säkerställd på signifikansnivåerna 5 respektive 1 procent.

En möjlig invändning mot jämförelsen mellan Tabell 6 och Tabell 5 är att det inte bara är typen av utfallsmått som skiljer utan också sätten att mäta dem. Medan percentilrang, som används i Tabell 5, ligger mycket nära en kontinuerlig variabel kan betygspoäng, som används i Tabell 6, endast anta fyra olika värden, nämligen 0, 10, 15 och 20. För att undersöka hur viktig denna invändning är kan vi utnyttja att resultaten på de nationella proven inte bara redovisas som en provpoäng utan också i form av ett provbetyg, mätt i samma enheter som kursbetyget.

Det visar sig att kvalitativt sett blir vår slutsats när vi mäter resultaten på det nationella provet i termer av provbetyg identisk med när vi mäter resultaten i termer av (percentilrangordnade) provpoäng. Detta framgår av Tabell 7 som dels innehåller den information som finns i Tabell 5 och Tabell 6, dels motsvarande information när utfallsmåttet utgörs av betyg på det nationella provet. För fysik motsvarar effekten i det senare fallet en knapp tredjedel av avståndet mellan betygsstegen G, VG och MVG.<sup>12</sup>

Slutsatsen blir alltså att den skillnad i skattade effekter av NTA som vi finner när vi använder (percentilrangordnade) provpoäng respektive kursbetyg *inte* beror på att de mäts utefter olika skalor. Vad orsaken är kan vi bara försöka

<sup>12</sup>  $1,56/5 \approx 0,31$ , där nämnaren (5) är lika med poängskillnaden (VG – G) resp. (MVG – VG).

gissa oss till. En sak att notera är då att kursbetyg är ett ”bredare” utfallsmått än provpoäng: resultatet på det nationella provet ska vara av betydelse för kursbetyget men kursbetyget ska även avspegla en mängd andra förhållanden, t.ex. resultat på andra prov, aktivitet på lektioner o.s.v. Ju högre vikt dessa andra förhållanden tilldelas desto svagare kommer sambandet mellan provpoäng och kursbetyg att bli. Och så länge andra faktorer än resultatet på det nationella provet viktas på olika sätt av olika lärare så kvarstår det grundläggande problemet med kursbetyg i detta sammanhang: de är inte lika jämförbara som resultaten på nationella prov.<sup>13</sup>

Tabell 7 Resultat på nationella prov – percentilrangordnade poäng och provbetyg – samt kursbetyg i biologi, fysik och kemi för NTA-deltagare och icke-NTA-deltagare som parats ihop (matchats) med hjälp av skattade propensity scores

|                         | NTA-deltagare    |                    |                   | Icke-NTA-deltagare |       |                   | NTA – icke-NTA |                   |
|-------------------------|------------------|--------------------|-------------------|--------------------|-------|-------------------|----------------|-------------------|
|                         | Antal individer  | Medel <sup>1</sup> | Standardavvikelse | Antal individer    | Medel | Standardavvikelse | Differens      | Standardavvikelse |
| Percentilrang provpoäng |                  |                    |                   |                    |       |                   |                |                   |
| <i>Biologi</i>          | 339              | 45,38              | 28,40             | 339                | 43,88 | 26,85             | 1,50           | 2,120             |
| <i>Fysik</i>            | 322              | 49,75              | 27,45             | 313                | 42,47 | 29,14             | 7,28**         | 2,242             |
| <i>Kemi</i>             | 339              | 50,18              | 28,15             | 348                | 46,60 | 29,23             | 3,58           | 2,187             |
| <i>Samtliga</i>         | 1 000            | 48,42              | 28,07             | 1 000              | 44,38 | 28,44             | 4,04**         | 1,263             |
| Provbetyg               |                  |                    |                   |                    |       |                   |                |                   |
| <i>Biologi</i>          | 339              | 10,15              | 5,729             | 339                | 10,13 | 5,293             | 0,02           | 0,423             |
| <i>Fysik</i>            | 322              | 10,73              | 4,899             | 313                | 9,17  | 5,797             | 1,56**         | 0,425             |
| <i>Kemi</i>             | 339              | 10,10              | 5,486             | 348                | 9,38  | 5,879             | 0,72           | 0,433             |
| <i>Samtliga</i>         | 1 000            | 10,32              | 5,392             | 1 000              | 9,57  | 5,670             | 0,75**         | 0,247             |
| Kursbetyg               |                  |                    |                   |                    |       |                   |                |                   |
| <i>Biologi</i>          | 339              | 12,17              | 5,053             | 335                | 11,87 | 4,646             | 0,30           | 0,373             |
| <i>Fysik</i>            | 321              | 11,95              | 4,351             | 312                | 11,94 | 5,120             | 0,01           | 0,377             |
| <i>Kemi</i>             | 335              | 12,04              | 4,886             | 343                | 11,94 | 4,937             | 0,10           | 0,377             |
| <i>Samtliga</i>         | 995 <sup>2</sup> | 12,06              | 4,776             | 990 <sup>2</sup>   | 11,91 | 4,895             | 0,15           | 0,217             |

<sup>1</sup> Vid beräkningen av betygsmedelvärden har betygen värderats enligt: IG = 0, G = 10, VG = 15, MVG = 20.

<sup>2</sup> Antalet elever med kursbetyg är något lägre än de 1 000 elever vardera som utgör de matchade NTA- och icke-NTA-deltagarna, p.g.a. att några elever inte uppfyllde kraven för att kunna få kursbetyg.

<sup>3</sup> Differens = medelvärdet för NTA-deltagare – medelvärdet för individer som ej deltagit i NTA. Symbolerna \* och \*\* markerar att ett dubbelsidigt t-test visar att differensen är statistiskt säkerställd på signifikansnivåerna 5 respektive 1 procent.

<sup>13</sup> Att olika lärare faktiskt tillmäter de nationella proven olika betydelse vid betygssättningen är högst troligt. Enligt Skolverket (2011, s.1): ”... i hur hög grad ett nationellt prov ska påverka betygssättningen finns inte reglerat. Det är läraren själv som bestämmer hur hon eller han ska tolka att det nationella provets resultat ska utgöra ett stöd vid betygssättningen.”

I Skolverket (2007) diskuteras ett antal tänkbara förklaringar till avvikelser mellan provbetyg och kursbetyg. Där nämns bland annat att läraren har att ta hänsyn till flera mål än de som ingår i det nationella provet och även har ett mer omfattande underlag för sin bedömning än elevens provresultat (jfr. ovan), att särskilda insatser kan ha satts in för elever som inte klarat proven bra samt att eleverna har arbetat med ett annat stoff än det provet behandlar.

Den sista förklaringen, att eleverna har arbetat med ett annat stoff än det provet behandlar, är av speciellt intresse i detta sammanhang. Skälet är att de nationella proven i de naturvetenskapliga ämnena testar två olika färdigheter: dels teori- och faktakunskaper, dels förmågan att planera och genomföra en vetenskaplig undersökning. Det är framför allt den senare färdigheten som deltagande i NTA kan förbättra.<sup>14</sup> Elever som har deltagit i NTA bör ha arbetat mer med, och ha större färdigheter i, att planera och genomföra vetenskapliga undersökningar än andra elever. Vid betygssättningen är det därför rimligt att lärare vars elever har deltagit i NTA fäster större vikt vid de resultat i de nationella proven som avser planering och genomförande av vetenskapliga undersökningar än vad andra lärare gör. Denna hypotetiska skillnad i betygssättningen har två konsekvenser som vi bör kunna observera i data.

Den första konsekvensen är att effekterna av NTA ska vara mera påtagliga när elevernas färdigheter mäts i termer av nationella prov än när de mäts i termer av kursbetyg. Detta har vi redan konstaterat.

Tabell 8 Skillnader mellan kursbetyg och provbetyg i årskurs 9 i biologi, fysik och kemi för matchade icke-NTA-deltagare och NTA-deltagare samt skillnaderna mellan dessa skillnader

| Ämne     | Icke-NTA-deltagare           |  |                   | NTA-deltagare   |                                   |                   | Icke-NTA – NTA                     |                   |
|----------|------------------------------|--|-------------------|-----------------|-----------------------------------|-------------------|------------------------------------|-------------------|
|          | Antal individer <sup>2</sup> | Medeldiff <sup>1</sup> (kursbetyg – provbetyg) | Standardavvikelse | Antal individer | Medeldiff (kursbetyg – provbetyg) | Standardavvikelse | Differens i medeldiff <sup>3</sup> | Standardavvikelse |
| Biologi  | 335                          | 1,74   | 3,906             | 339             | 2,02                              | 4,067             | -0,29                              | 0,307             |
| Fysik    | 312                          | 2,77   | 4,384             | 321             | 1,25                              | 3,331             | 1,52**                             | 0,309             |
| Kemi     | 343                          | 2,48   | 4,273             | 335             | 1,93                              | 3,939             | 0,55*                              | 0,316             |
| Samtliga | 990                          | 2,32   | 4,207             | 995             | 1,74                              | 3,812             | 0,58**                             | 0,180             |

<sup>1</sup> Vid beräkningen av betygsmedelvärden har betygen värderats enligt: IG = 0, G = 10, VG = 15, MVG = 20.

<sup>2</sup> Antalet elever med kursbetyg är något lägre än de 1 000 elever vardera som utgör de matchade NTA- och icke-NTA-deltagarna, p.g.a. att några elever inte uppfyllde kraven för att kunna få kursbetyg.

<sup>3</sup> Differens i medeldiff = medeldiff (kursbetyg – provbetyg) för icke-NTA-deltagare – medeldiff (kursbetyg – provbetyg) för NTA-deltagare. Symbolerna \* och \*\* markerar att ett *enkelsidigt t*-test visar att differensen är statistiskt säkerställd på signifikansnivåerna 5 respektive 1 procent.

<sup>14</sup> Detta följer av NTA:s experimentbaserade utformning och styrks även av NTA-representanter.

Den andra konsekvensen är att skillnaderna mellan provbetyg och kursbetyg ska vara mindre bland de elever som har deltagit i NTA än de elever som inte har deltagit. Om vi beräknar den genomsnittliga skillnaden mellan kursbetyg och provbetyg för icke-deltagare i NTA och minskar den med den genomsnittliga skillnaden mellan kursbetyg och provbetyg bland NTA-deltagare bör vi således få ett positivt tal. Sådana beräkningar redovisas i Tabell 8. Det visar sig att i tre fall av fyra är skillnaden positiv, som förväntat, nämligen för fysik, kemi och för de tre ämnena sammantagna, men inte för biologi. De tre positiva skillnaderna är också signifikant större än noll. Tabell 8 ger alltså också stöd för hypotesen att lärare vars elever har deltagit i NTA i sin betygsättning fäster större vikt vid de resultat i de nationella proven som avser planering och genomförande av vetenskapliga undersökningar än vad andra lärare gör.

## 6 Sammanfattning och diskussion

I denna artikel har vi gått från det partikulära till det generella. Vi har visat hur resultaten från en utvärdering av effekterna av en speciell utbildningsinsats – NTA – kan användas för att diskutera tre olika aspekter som är relevanta för effektutvärderingar av utbildningsinsatser i största allmänhet.

Den första aspekten avser vikten av att konstruera en rättvisande jämförelsegrupp. Vi har visat att om man bortser från denna aspekt kan analysens slutsatser bli felaktiga inte bara ifråga om effekternas storlek utan även ifråga om deras riktning. Effekten av NTA på nationella prov i naturvetenskap i årskurs 9 framstod som signifikant negativ när systematiska skillnader mellan NTA-deltagare och icke-deltagare inte beaktades. Vid kontroll för skillnaderna mellan grupperna blev effekten istället signifikant positiv.

Den andra aspekten har att göra med det välkända faktumet att en och samma medicin kan vara olika verksam beroende på mot vilka åkommor den används. Vår analys visade att NTA har en statistiskt säkerställd positiv inverkan på resultaten i fysik men inte i biologi och kemi.

Den tredje aspekten visar på betydelsen av att effekterna mäts i förhållande till ett utfallsmått som är standardiserat och därmed direkt jämförbart mellan olika elever och olika skolor. Resultaten på de nationella proven uppfyller detta villkor – frågorna är desamma för alla elever och svaren rättas efter en enhetlig mall. Kursbetygen uppfyller inte villkoret – de baseras på flera andra prestationer än resultaten på de nationella proven och vilka dessa andra faktorer är och hur de vägs ihop skiljer sig mellan olika lärare och skolor, och möjligen även mellan elever. Liksom ifråga om den första aspekten utgör vår analys ett

exempel på att det verkligen spelar roll hur man gör. När resultat på nationella prov används som utfallsvariabel finner vi signifikanta effekter av NTA. Med kursbetyg är skillnaderna mellan NTA-deltagare och elever som inte deltagit i NTA mycket små och inte i närheten av att vara statistiskt säkerställda. Vi finner stöd för att lärare vars elever har deltagit i NTA tenderar att sätta kursbetyg som ligger närmare resultaten på de nationella proven än vad kursbetygen gör för de elever som inte har deltagit i NTA. Vår hypotetiska förklaring till detta resultat är att lärare vars elever inte har deltagit i NTA i sin betygssättning fäster lägre avseende vi den del av de nationella proven som avser planering och genomförande av vetenskapliga undersökningar än vad lärare med NTA-deltagare som elever gör. Vi finner också stöd för denna hypotetiska förklaring i våra resultat.

Utöver dessa tre aspekter på effektutvärderingar av utbildningsåtgärder finns det naturligtvis ytterligare distinktioner som är av allmänt intresse. En som vi indirekt har berört är skillnaden mellan kvalitativa och kvantitativa analyser. Avsnitt 3 ger vid handen att de kvalitativa analyserna har resulterat i en mera positiv bedömning av NTA än vår kvantitativa studie har gjort. Men resultaten behöver inte stå i motsättning till varandra. Det är inte omöjligt att de kvalitativa studierna har avsett lärare och/eller elever vars erfarenheter av NTA av olika skäl har varit ovanligt positiva.<sup>15</sup> I så fall kan studiernas representativitet ifrågasättas men inte nödvändigtvis deras resultat.<sup>16</sup> Dessutom har vi i avsnitt 4 pekat på två skäl till att vår analys tenderar att underskatta positiva effekter av NTA.<sup>17</sup> Skillnaden i resultat mellan de tidigare kvalitativa studierna och vår kvantitativa studie är alltså kanske i realiteten mindre än de i förstone verkar vara.

I detta sammanhang vill vi påpeka att det faktum att den kvantitativa metod som ligger till grund för denna artikel är Propensity Score Matching (PSM) inte innebär att vi menar att PSM är en bra utvärderingsmetod i största allmänhet. När så är möjligt, är det alltid bättre att planera och genomföra en utbildningsåtgärd på ett sådant sätt att åtgärden kan utvärderas som om den genomförts på samma sätt som ett (medicinskt) experiment. Det behöver inte innebära att man lottar om att vissa ska få delta och andra inte. Det kan mycket väl vara så att

---

<sup>15</sup> Det skulle till exempel kunna bero på att det är lättare att höra talas om skolor där man framgångsrikt har tillämpat NTA än skolor där användningen av NTA har fungerat mindre bra. På motsvarande sätt kan det vara lättare att få samtycke till intervjuer och frågor om NTA i skolor där lärare och elever är positiva till NTA.

<sup>16</sup> Annorlunda uttryckt: analysernas *externa validitet* kan ifrågasättas men deras *interna validitet* kan vara god.

<sup>17</sup> Vi visade att både bortfall i stickprovsundersökningen av elever som skrev de nationella proven och fel i klassificeringen av elever som deltagare respektive icke deltagare i NTA medför att positiva effekter av NTA underskattas.

alla får delta men att man lottar om *när* olika individer ska påbörja sitt deltagande, se vidare Mellander (2015).

Skälet till att ett experimentliknande upplägg är att föredra är att det medför mycket lägre krav på den information som krävs för utvärderingen än vad PSA gör. Med experimentupplägget kan man som mått på åtgärdens effekt helt enkelt beräkna skillnaden mellan det genomsnittliga utfallet (resultatet) i behandlingsgruppen (deltagarna) med det genomsnittliga utfallet i kontrollgruppen (icke-deltagarna, alternativt de som deltagit senare). Skälet är att genom lottningen av deltagare kan individerna i behandlings- och kontrollgruppen förväntas vara jämförbara både vad gäller observerbara och icke-observerbara egenskaper. Som nämnts i avsnitt 5.1, kräver PSM däremot att utvärderaren också kan definiera och mäta samtliga faktorer som har påverkat huruvida en individ har deltagit. Detta stränga villkor beror på att utvärderaren inte har haft kontroll över hur deltagandet i åtgärden har bestämts.

Vad gäller framtida forskning bör det finnas goda möjligheter att kombinera kvantitativa och kvalitativa ansatser. En studie där man först genomför en kvantitativ analys och sedan följer upp denna med kvalitativa undersökningar av några skolor/elever/lärare har tre viktiga fördelar. Den första är att den kvantitativa analysen kan användas som urvalsinstrument för den kvalitativa analysen. Man vet då hur de enheter man väljer att studera närmare förhåller sig till den representativa enheten, den som beskrivs av genomsnittresultaten. Den andra fördelen är att man genom att närstudera enheter vars resultat överensstämmer med den representativa enhetens kan få information om vilka mekanismer som har genererat genomsnittresultaten. Den tredje fördelen är att fallstudier av enheter vars resultat tydligt avviker från den representativa enhetens kan ge information om vad som karakteriserar enheter som ligger långt ifrån genomsnittet.



## Referenser

- Anderhag, P. & Wickman, P. – O. (2006), ”NTA som kompetensutveckling för lärare. Utvärdering av hur lärares deltagande i NTA utvecklar deras kompetens att stödja elevernas begrepps- och språkutveckling”, Stockholm: Lärarhögskolan, UKL.
- Anderhag, P. & Wickman, P. – O. (2007), ”Utvärdering av hur NTA hjälper skolorna att nå kursplanemålen för femte skolåret i naturorienterande ämnen”, Stockholm: Lärarhögskolan, UKL.
- Geier, R. et al., “Standardized test outcomes, for students engaged in inquiry-based science curricula in the context of urban reform”, *Journal of Research in Science Teaching*, 45(8), 922–939.
- Gisselberg, K. (2001), ”NTA-projektets tre första år – en positionsbestämning”, Umeå universitet, Lärarutbildningen, IMN.
- Guo, S. & Fraser, M.W. (2010), *Propensity Score Analysis: Statistical methods and applications*, Advanced Quantitative Techniques in the Social Sciences 11. Los Angeles: Sage Publications.
- Mellander, E. (2015), “Möjligheterna att bedöma, följa upp och effektutvärdera utbildnings- och fortbildningspolitiken”, *Nordic Studies in Education*, No 4.
- Mellander, E. & Svärth, J. (2015), ”Inquiry-based learning put to test: Long-term effects of the Swedish Science and Technology for Children program”, IFAU Working Paper 2015:23
- Schoultz, J. & Hultman, G. (2002), ”Det är bra med NTA. Vi gör inte saker för att tråka ut oss utan för att lära oss. Utvärdering av elevers och lärares lärande och utveckling inom NTA-projektet”, Linköpings universitet, ITUF, IUUV.
- Schoultz, J., Hultman, G. & Lindkvist, M. (2003), ”I början fick vi använda vår fantasi. Utvärdering av elevers och lärares lärande och utveckling inom NTA-projektet”, Linköpings universitet, IUUV.
- Skolverket (2007). ”Provbetyg – Slutbetyg – Likvärdig bedömning? En statistisk analys av sambandet mellan nationella prov och slutbetyg i grundskolans årskurs 9, 1998–2006”, Rapport 300.
- Skolverket (2011), ”Skillnaden mellan betygsresultat på nationella prov och ämnesbetyg i årskurs 9, läsåret 2009/10”, PM 71-2011-14.

## Appendix

Tabell A1 Logistisk regression för beräkning av propensity scores. Beroende variabel 1 för NTA-individer och 0 för icke-NTA-individer

| Variabel  | Parameter-skattning  | Standard-fel | Signifikans-nivå |
|---|----------------------|--------------|------------------|
| Skolnivå <sup>1</sup> :   |                      |              |                  |
| Andel pojkar  | 4,565** <sup>9</sup> | 0,731        | 0,000            |
| Föräldrarnas utbildningsnivå <sup>2</sup>                                     | - 0,833**            | 0,310        | 0,007            |
| Andel utlandsfödda elever   | - 3,269**            | 0,711        | 0,000            |
| Andel elever med utlandsfödda föräldrar                                       | 3,641**              | 0,616        | 0,000            |
| Andel elever i åk 9 med minst Godkänt i alla ämnen                            | - 1,946**            | 0,516        | 0,000            |
| Kommunnivå <sup>3</sup> :   |                      |              |                  |
| Befolkning, i miljoner  | - 4,921**            | 0,342        | 0,000            |
| Andel av befolkningen i tätbebyggda <sup>4</sup> områden                      | 2,865**              | 0,447        | 0,000            |
| Medianinkomst <sup>5</sup> , miljoner SEK                                     | 52,223**             | 3,105        | 0,000            |
| Andel lärare med pedagogisk universitetsutbildning <sup>6</sup>               | 9,253                | 0,869        | 0,000            |
| Antal elever per 100 lärare   | 19,742               | 6,515        | 0,002            |
| Totala kostnader <sup>7</sup> för åk 1–9 per elev i åk 1-9 som bor i kommunen | 114,351**            | 8,318        | 0,000            |
| Andel elever i åk 9 med minst Godkänt i alla ämnen                            | - 0,112**            | 0,010        | 0,000            |
| Länsnivå:   |                      |              |                  |
| Länsdummy-variabler <sup>8</sup>  | Ja                   |              |                  |

1. Skolnivåvariablerna utgör genomsnitt för åren 2004–06.
2. Föräldrarnas utbildning är kodad som 1 för grundskola, 2 för gymnasium och 3 om utbildningen innehåller minst en termins universitetsstudier. Variabeln är beräknad som genomsnittet av utbildningskoderna bland föräldrarna till eleverna på skolan.
3. Kommunivåvariablerna avser år 2006.
4. Områden med minst 200 invånare där avståndet mellan bostäderna är högst 200 m.
5. Medianinkomsten, efter skatter och transfereringar, bland 20+ år gamla invånare.
6. I pedagogiska universitetsutbildningar ingår förskollärare och fritidspedagoger. Antalet är räknat i heltidsekvivalenter.
7. Kostnaderna är räknade netto, efter statsbidrag.
8. För att spara på utrymme rapporteras inte de skattade koefficienterna för länsdummy-variablerna.
9. Symbolerna \* och \*\* markerar att ett *t*-test visar att parameterskattningen är signifikant skild ifrån noll på signifikansnivåerna 5 respektive 1 procent.

## IFAU:s publikationsserier – senast utgivna

### Rapporter

- 2015:1** Albrecht James, Peter Skogman Thoursie och Susan Vroman ”Glastaket och föräldraförsäkringen i Sverige”
- 2015:2** Persson Petra ”Socialförsäkringar och äktenskapsbeslut”
- 2015:3** Frostenson Magnus ”Organisatoriska åtgärder på skolnivå till följd av lärarlegitimationsreformen”
- 2015:4** Grönqvist Erik och Erik Lindqvist ”Kan man lära sig ledarskap? Befälsutbildning under värnplikten och utfall på arbetsmarknaden”
- 2015:5** Böhlmark Anders, Helena Holmlund och Mikael Lindahl ”Skolsegregation och skolval”
- 2015:6** Håkanson Christina, Erik Lindqvist och Jonas Vlachos ”Sortering av arbetskraftens förmågor i Sverige 1986–2008”
- 2015:7** Wahlström Ninni och Daniel Sundberg ”En teoribaserad utvärdering av läroplanen Lgr 11”
- 2015:8** Björvang Carl och Katarina Galic’ ”Kommunernas styrning av skolan – skolplaner under 20 år”
- 2015:9** Nybom Martin och Jan Stuhler ”Att skatta intergenerationella inkomstsamband: en jämförelse av de vanligaste måtten”
- 2015:10** Eriksson Stefan och Karolina Stadin ”Hur påverkar förändringar i produkt efterfrågan, arbetsutbud och lönekostnader antalet nyanställningar?”
- 2015:11** Grönqvist Hans, Caroline Hall, Jonas Vlachos och Olof Åslund ”Utbildning och brottslighet – vad händer när man förlängde yrkesutbildningarna på gymnasiet?”
- 2015:12** Lind Patrik och Alexander Westerberg ”Yrkeshögskolan – vilka söker, vem tar examen och hur går det sedan?”
- 2015:13** Mörk Eva, Anna Sjögren och Helena Svaleryd ”Hellre rik och frisk – om familjebakgrund och barns hälsa”
- 2015:14** Eliason Marcus och Martin Nilsson ”Inläsningseffekter och differentierade ersättningsnivåer i sjukförsäkringen”
- 2015:15** Boye Katarina ”Mer vab, lägre lön? Uttag av tillfällig föräldrapenning för vård av barn och lön bland svenska föräldrar”
- 2015:16** Öhman Mattias ”Smarta och sociala lever längre: sambanden mellan intelligens, social förmåga och mortalitet”
- 2015:17** Mellander Erik och Joakim Svärth ”Tre lärdomar från en effektutvärdering av lärarstödsprogrammet NTA”

## Working papers

- 2015:1** Avdic Daniel “A matter of life and death? Hospital distance and quality of care: evidence from emergency hospital closures and myocardial infarctions”
- 2015:2** Eliason Marcus “Alcohol-related morbidity and mortality following involuntary job loss”
- 2015:3** Pingel Ronnie och Ingeborg Waernbaum “Correlation and efficiency of propensity score-based estimators for average causal effects”
- 2015:4** Albrecht James, Peter Skogman Thoursie och Susan Vroman “Parental leave and the glass ceiling in Sweden”
- 2015:5** Vikström Johan “Evaluation of sequences of treatments with application to active labor market policies”
- 2015:6** Persson Petra “Social insurance and the marriage market”
- 2015:7** Grönqvist Erik och Erik Lindqvist “The making of a manager: evidence from military officer training”
- 2015:8** Böhlmark Anders, Helena Holmlund och Mikael Lindahl “School choice and segregation: evidence from Sweden”
- 2015:9** Håkanson Christina, Erik Lindqvist och Jonas Vlachos “Firms and skills: the evolution of worker sorting”
- 2015:10** van den Berg Gerard J., Antoine Bozio och Mónica Costa Dias “Policy discontinuity and duration outcomes”
- 2015:11** Wahlström Ninni och Daniel Sundberg “Theory-based evaluation of the curriculum Lgr 11”
- 2015:12** Frölich Markus och Martin Huber “Direct and indirect treatment effects: causal chains and mediation analysis with instrumental variables”
- 2015:13** Nybom Martin och Jan Stuhler “Biases in standard measures of inter-generational income dependence”
- 2015:14** Eriksson Stefan och Karolina Stadin “What are the determinants of hiring? – The role of demand and supply factors”
- 2015:15** Åslund Olof, Hans Grönqvist, Caroline Hall och Jonas Vlachos “Education and criminal behaviour: insights from an expansion of upper secondary school”
- 2015:16** van den Berg Gerard J. och Bas van der Klaauw “Structural empirical evaluation of job search monitoring”
- 2015:17** Nilsson Martin “Economic incentives and long-term sickness absence: the indirect effect of replacement rates on absence behaviour”
- 2015:18** Boye Katarina “Care more, earn less? The association between care leave for sick children and wage among Swedish parents”

- 2015:19** Assadi Anahita och Martin Lundin “Tenure and street level bureaucrats: how assessment tools are used at the frontline of the public sector”
- 2015:20** Stadin Karolina “Firms’ employment dynamics and the state of the labor market”
- 2015:21** Öhman Mattias “Be smart, live long: the relationship between cognitive and non-cognitive abilities and mortality”
- 2015:22** Hägglund Pathric, Per Johansson och Lisa Laun “Rehabilitation of mental illness and chronic pain – the impact on sick leave and health”
- 2015:23** Mellander Erik och Joakim Svårdh “Inquiry-based learning put to test: long-term effects of the Swedish science and technology for children program”

### **Dissertation series**

- 2014:1** Avdic Daniel “Microeconomic analyses of individual behaviour in public welfare systems”
- 2014:2** Karimi Arizo “Impacts of policies, peers and parenthood on labor market outcomes”
- 2014:3** Eliasson Tove “Empirical essays on wage setting and immigrant labor market opportunities”
- 2014:4** Nilsson Martin “Essays on health shocks and social insurance”
- 2014:5** Pingel Ronnie “Some aspects of propensity score-based estimators for causal inference”
- 2014:6** Karbownik Krzysztof “Essays in education and family economics”