

TEMA: Storskaliga studier

Kunskapsmätningar som paradox och möjlighet

Anders Jakobsson

Ett övergripande fokus i detta temanummer är att analysera, granska och diskutera de storskaliga studierna och deras påverkan på utbildningssystemen i Sverige och internationellt. I flera av artiklarna argumenterar författarna för att resultaten från PISA och TIMSS såväl som de nationella proven tenderar att få ett allt större inflytande på utbildningspolitik och undervisning i Sverige och i övriga världen under 2000-talet. Ett sådant exempel är de senaste årens tillkomst av ett antal skolreformer, såsom nya läroplaner, betygssystem, skollag och en ny (-gammal) lärarutbildning. Vid en granskning av de politiska dokumenten bakom besluten är det möjligt att upptäcka att man vid upprepade tillfällen använder resultatbeskrivningar från de storskaliga studierna som argument för att genomföra reformerna (se t.ex. SOU 2008:109). Detta trots att det är högst tveksamt om man kan dra denna typ av långtgående slutsatser utifrån resultaten från PISA och TIMSS. Tendensen att använda de storskaliga studierna i den politiska retoriken och som underlag för beslutsfattande är inte unikt för Sverige. Exempelvis visar Sotiria Grek (2009) i en intervjustudie av europeiska skolpolitiker att nationsrankingar blivit en alltmer inflytelserik påverkansfaktor för en ny skolpolitisk strategi när det gäller att styra utbildningssystem i Europa. Det fungerar genom att politiker utnyttjar internationella resultatbeskrivningar från OECD och IEA för att legitimera sina egna förslag om förändringar utifrån argumentet att testerna utgör entydiga och robusta bevis.

Frågan blir då vilka slutsatser man egentligen kan dra från denna typ av studier och hur det vetenskapliga samhället uppfattar dem? Hur

Anders Jakobsson är professor i naturvetenskapernas didaktik vid fakulteten för Lärande och Samhälle vid Malmö Högskola. Malmö högskola, Lärande och samhälle, 20506 Malmö. E-post: anders.jakobsson@mah.se

pass tillförlitliga är resultaten och vad kan man säga om studiernas validitet och reliabilitet? Denna typ av frågor leder naturligtvis till ett antal olika svar beroende på vilka utbildningsvetenskapliga paradigman man undersöker (se t.ex. Biesta 2010, Jakobsson et al 2013, Mayer & Benavot 2013). Emellertid visar Harvey Goldstein (2004) i en analys av PISA-testernas reliabilitet och statistiska metodologi att tvärsnittsdata (cross-sectional data) som används för att genomföra nationsrankingen i PISA, i princip, gör det omöjligt att dra tillförlitliga slutsatser om de deltagande ländernas skolsystem. Även Michael Fertig (2004) framhåller att man endast kan dra slutsatser från nationsrankingen om man utgår från starka antaganden om att det existerar stora likheter i ländernas utbildningssystem och Peter Allrup (2007) menar att de statistiska skalorna i PISA inte är homogena med tanke på ett antal bakgrundsvariabler vilket leder till en misstolkning av Rasch-modellen vilken i sin tur utgör själva utgångspunkten för metodologin bakom nationsrankingarna. Dessa studier och flera andra indikerar att det existerar en tveksamhet från det vetenskapliga samhället när det gäller att dra alltför långtgående slutsatser från studierna. Framförallt riktar sig kritiken mot nationsrankingarna mot tendensen att presentera resultaten i form av endimensionella ”League Tables” över de deltagande ländernas resultat och mot att använda nationsrankingarna som ett förenklat index över kvaliteten i nationers skol- och utbildningssystem. Denna kritik är med andra ord huvudsakligen inriktad på den psykometriska konstruktionen av de nationella medelvärdena och de nationella rankingarna.

På den andra sidan i debatten om de storskaliga studiernas tillförlitlighet står OECD och IEA som naturligtvis argumenterar för att studierna har hög grad av reliabilitet och validitet. I en text om ramverket för PISA studierna påstår OECD följande:

PISA is the most comprehensive and rigorous international programme to assess student performance and to collect data on the student, family and institutional factors that can help to explain differences in performance. Decisions about the scope and nature of the assessments and the background information to be collected are made by leading experts in participating countries, and are steered jointly by governments on the basis of shared, policy-driven interests. Substantial efforts and resources are devoted to achieving cultural and linguistic breadth and balance in the assessment materials. Stringent quality assurance mechanisms are applied in translation, sampling and data collection. As a consequence, the results of PISA have a high degree of validity and reliability, and can significantly improve understanding of the outcomes of education in the world's economically most developed countries, as well as in a growing number of countries at earlier stages of economic development (OECD 2009, s 13).

I texten framhålls fördelarna med PISA-tester och de beskrivs som om de relativt entydigt och med hög grad av tillförlitlighet kan beskriva utbildningssystemens effektivitet i de ekonomiskt mest utvecklade länderna i världen såväl som länder i tidigare stadier av ekonomisk utveckling. Ingenstans i texten diskuteras den framförda kritiken mot nationsrankingar eller problematiken när det gäller att utvärdera kvaliteten i olika länders skolsystem. Texten saknar dessutom en generell diskussion om tillförlitligheten i metodologin bakom de internationella rankingarna. Trots dessa brister tycks det samtidigt finnas en tendens att efterfrågan från det politiska samhället på denna typ av produkter ständigt ökar. Det är också sannolikt att anta att internationella kunskapsmätningar och olika typer av kunskapsindex kommer att få ett ökat inflytande över de nationella skolsystemen. Ett exempel på denna utveckling är att OECD från och med 2015 överlåter delar av uppdraget att genomföra PISA-testerna till utvärderingsföretaget *Pearson* som är världens största företag inom utbildning och utvärdering. Företaget beskriver i sitt program *The learning curve* att man avser att skapa ett världsindex ”of cognitive skills and educational attainment” som baseras på resultaten från internationella kunskapsmätningar, utvärdering av medborgares allmänbildning och statistik om befolkningens utbildningsnivå (se <http://thelearningcurve.pearson.com/>). Förhoppningen är att skapa ett globalt nyckeltal som på ett enkelt och enhetligt sätt kan beskriva de deltagande ländernas utbildningssystem och därmed också utgöra ett viktigt beslutsunderlag för det utbildningspolitiska samhället.

Mot denna bakgrund har det i detta temanummer varit ett viktigt syfte att granska, analysera och diskutera de internationella kunskapsmätningarna utifrån tillförlitlighet och användning. Flera av författarna framför kritik mot resultatbeskrivningar i form av nationella rankingar och mot att resultaten används som förenklade nyckeltal och beskrivningar över utbildningsvetenskapliga frågor av komplext innehåll. Dessutom diskuterar några av författarna i vilken utsträckning denna typ av tester förmår att utgöra valida representationer av elevernas faktiska kunskaper och kompetenser i ett modernt samhälle utifrån en sociokulturell teoriram. Samtidigt framhåller flera författare vikten av att det utbildningsvetenskapliga samhället inte ställer sig utanför den pågående debatten om de storskaliga studierna utan att de istället deltar i arbetet med att granska värdet av dem. Gemensamt för artiklarna i detta temanummer är att författarna genom att ta sin utgångspunkt i olika typer av data från de storskaliga studierna därmed också bidrar till att granska tillförlitligheten i dem. Exempelvis diskuterar *Anders Jakobsson* i ATT UNDERSÖKA KUNSKAPSTRENDER MED HJÄLP AV PISA – LIKVÄRDIGHET,

FÖRSTÅELSE OCH KUNSKAPSSYN hur man kan återanvända data från de storskaliga studierna för att undersöka trender när det gäller svenska elevers kunskaper i och om naturvetenskap under 2000-talet. Syftet med artikeln är framförallt att beskriva alternativa eller komplementära analyser av PISA-data för att kunna relatera kunskapstrender till faktorer som likvärdighet, kunskapssyn och elevers förståelse av ämnesområdet. Det handlar med andra ord inte om att beskriva enskilda elevers kunskapsnivå eller en allmän nationell nivå utan mer att analysera generella strukturer och underliggande mönster i resultaten. Analyserna visar exempelvis att den nedåtgående trenden i de svenska resultaten i naturvetenskap och läsning huvudsakligen kan förklaras utifrån att det är de lågpresterande och i viss mån de medelpresterande elevernas resultat som kraftigt försämrats under 2000-talet. Dessutom visar analyserna att resultatskillnaden uttryckt på skolnivå har fördubblats i naturvetenskap och nästan tredubblats i läsning vilket indikerar en kraftig ökning av kunskapssegregationen i svensk skola under perioden 2000-2009. Vi har alltså fått en tydlig kunskapsuppdelning och polarisering mellan hög- och lågpresterande skolor i Sverige under denna period. Fördjupade analyser med fokus på testernas kunskapsinnehåll visar att de svenska eleverna generellt har förbättrat sina resultat på faktabaserade uppgifter och på uppgifter som mäter enkla ämneskunskaper trots att de generella resultaten visar en tydligt nedåtgående trend. En förklaring till denna tydliga trend är att elevernas resultat på uppgifter som avser att mäta problemlösning, kritiskt tänkande och frågeställningar som belyser naturvetenskapens karaktär sjunkit kraftigt under 2000-talet.

Eva Davidsson, Karl-Göran Karlsson och Magnus Oskarsson använder i sin artikel, TRENDER OCH LIKVÄRDIGHET – SVENSKA ELEVERS RESULTAT PÅ PISA NATURVETENSKAP I EN INTERNATIONELL JÄMFÖRELSE, också trendanalyser för att jämföra de svenska PISA-resultaten i naturvetenskap med utvecklingen i andra europeiska länder. I artikeln konstaterar författarna att det existerar stora likheter i resultatutvecklingen för Sverige, Island och Tjeckien vilket innebär att det framförallt är de låg- och medelpresterande eleverna som försämrar sina resultat, att mellanskolevariansen (resultaten uttryckt per skolenhet) ökar och att likvärdigheten i ländernas skolsystem minskar. Samtidigt är det möjligt att upptäcka att de danska och polska resultaten visar en tydlig förbättring för motsvarande grupper. En viktig slutsats är att det tycks finnas ett samband mellan en ökning av mellanskolevariansen och en nedåtgående trend för lågpresterande elever vilket i sin tur tycks resultera i en nedåtgående trend i de nationella resultaten. Författarna framhåller att en möjlig förklaring är att det fria skolvalet resulterat i en förstärkt skolsegregation genom att elever

väljer skolor beroende på ambition och social tillhörighet. De polska resultaten uppvisar däremot en motsatt utveckling med en minskad mellanskolevarians och förbättrade nationella resultat. En förklaring är att de polska skolmyndigheterna under slutet av 1990-talet genomförde en sammanhållen grundskola med inspiration från de nordiska utbildningssystemen. Emellertid är det intressant att mellanskolevariansen i Polen i sig inte tycks vara den faktor som är kraftigast relaterad till resultatutvecklingen, utan snarare den successiva och signifikanta minskningen av resultatskillnaderna mellan skolorna.

I den tredje artikeln, HÖGSTADIELEVERS ANVÄNDNING AV NATURVETENSKAPLIGT SPRÅKBRUK I KEMIÄMNET I TIMMS, analyserar och jämför *Birgitta Frändberg* och *Sofie Johansson-Kokkinakis* svenska elevsvar i TIMSS 2007 med läroböcker i Kemi för årskurs 8. Analysen fokuserar elevernas aktiva ordförråd och användning av språket när de skriver egenformulerade svar i provsituationen. Studien indikerar att det existerar ett tydligt samband mellan högpresterande elevers språkanvändning och kemiläroböckernas karakteristiska språkbruk och uttryckssätt. Detta innebär att högpresterande elever tenderar att använda sig av en större andel tekniska termer, fler akademiska ord, längre svar, fler passiver och fler vardagliga ord med ämnesspecifik betydelse än vad de lågpresterande eleverna gör. När det gäller att använda fler vardagliga ord med ämnesspecifik betydelse framhåller författarna att detta kan tolkas som ett tecken på ett djup i ordförrådet då eleven kan uppvisa kunskap om flera betydelser av ett ord. Vidare framkommer i en analys av språkanvändningen i olika prestationsgrupper att elever som inte når upp till den elementära nivån i TIMSS har tredubblats under åren 1995 till 2007 (från 3% till 9%). En viktig slutsats i artikeln är att ett ökat intresse för hur sociala och kommunikativa faktorer påverkar elevers lärande också bör fokusera språkets betydelse i naturvetenskapliga läroprocesser.

Ett annat exempel hur man kan re-analysera data från TIMSS är Marie Wibergs och Ewa Rolfsmans SCHOOL EFFECTIVENESS IN SCIENCE IN SWEDEN AND NORWAY VIEWED FROM A TIMMS PERSPECTIVE. Syftet med studien är framför allt att identifiera ett antal faktorer på skolnivå (school-level factors) och att undersöka eventuella samband mellan dessa faktorer och elevernas faktiska prestationer. Förhoppningen är också att analysen (multilevel) kan öka förståelsen av vilka faktorer som förklarar nedgången i svenska och norska elevers prestationer under perioden. Resultaten av analysen indikerar emellertid att det existerar relativt stora resultatskillnader mellan skolor som för-definieras som effektiva i jämförelse med de skolor som definierats som mindre effektiva vilket innebär att nedgången är relativt större i de effektiva skolorna. Denna slutsats är i högre utsträckning giltigt för

de svenska resultaten än för de norska. Samtidigt visar analysen att det endast är några få faktorer på skolnivå som är möjliga att relatera till elevernas faktiska prestationer och att dessa faktorer är olika i de två länderna. Exempelvis indikerar analysen att faktorn socioekonomisk status på skolnivå tycks vara relativt starkt relaterad till prestationer i Sverige men att motsvarande relation inte är signifikant i Norge. Analysen indikerar däremot att faktorer på elevnivå har ett bättre förklaringsvärde när det gäller beskriva skoleffektivitet. Som slutsats framhåller författarna att en viktig utmaning inför framtiden är att fortsätta att utveckla metoder för att kunna beskriva och förstå vilka faktorer på skolnivå som kan förklara resultaten av de storskaliga studierna. I den femte artikeln, NATIONELLA PROV I NO OCH LÄRARES VAL AV UNDERVISNING SINNEHÅLL, undersöker Eva Lundqvist och Malena Lidar hur införandet av nationella prov i naturvetenskap i Sverige har påverkat lärares innehållsval inom ämnesområdet. Författarna har med andra ord studerat vilket innehåll lärare talar om som nytt i jämförelse med hur de uppfattar sin egen undervisning före införandet av proven. Analysen visar att man kan identifiera tre teman som nytt innehåll i proven; naturvetenskaplig argumentation, vetenskapshistoria och laborationer. I diskussionsavsnittet framhåller författarna också att de nationella proven, så som de varit utformade 2009-2011, kan inrymma en potential för ämnesdidaktisk utveckling och att detta framförallt är relaterat till det ämnesinnehåll som behandlar vetenskapernas intellektuella processer. I analysen av lärarintervjuerna framkommer dessutom att det finns variationer i vad lärarna tolkar som viktigt innehåll och att lärarna upplever att innehållet i proven väl motsvarar den undervisning de själva bedriver. En möjlig förklaring kan vara att proven är så mångfacetterade att alla lärare kan hitta olika innehåll som överensstämmer med hur de själva upplever sin naturvetenskapliga undervisning. Samtidigt menar författarna att det kan finnas en risk att lärarna genom proven får en bekräftelse på att de har gjort rimliga innehållsval i sin undervisning och bedömningspraktik, vilket i sin tur skulle kunna innebära att de inte behöver förändra sin undervisning i någon större utsträckning. I den avslutande artikeln, ATT FÖRSTÅ NATURVETENSKAPLIGA KUNSKAPSMÄTNINGAR UTIFRÅN EN SOCIOKULTURELL TEORIAM, analyserar Margareta Serder 70 elevers kollaborativa arbete med några PISA-uppgifter utifrån en sociokulturell teoriam. Syftet med artikeln är framförallt att öka förståelsen för hur elever uppfattar och tar sig an de provuppgifter i naturvetenskap som används för att mäta deras kunskaper inom ämnesområdet. I analysen blir det tydligt att det existerar ett avstånd mellan vad PISA-testerna avser att mäta och vad eleverna faktiskt diskuterar och fokuserar under problemlösningssprocessen. I

studien framträder exempelvis elever som gör ett tydligt motstånd mot provets sätt att beskriva naturvetenskapliga kunskaper och elever som går vilse i de många möjliga kontexter som ett prov med intentionen att placera kunnandet i ett vardagligt sammanhang aktiverar. Andra elever tycks inte vilja privilegiera naturvetenskaplig kunskap framför annan kunskap eller bemästrar inte det akademiska språkbruk som används i testerna. Ett exempel på detta är en uppgift om växthuseffekten och den globala uppvärmningen där fiktiva elever i ungefär samma ålder som de verkliga eleverna agerar som naturvetare när de analyserar vetenskapliga diagram gällande relationen mellan atmosfärens medeltemperatur och koldioxidutsläppen. I den fiktiva diskussionen använder eleverna ett avancerat naturvetenskapligt språkbruk för att generera olika hypoteser och förklaringsmodeller. Analysen visar att en stor del av de verkliga eleverna framförallt uttrycker ett tydligt avståndstagande till de fiktiva elevernas språkbruk och sätt att agera samt att detta utgör ett hinder att engagera sig i uppgiftens kunskapsinnehåll. Som slutsats framhåller författaren risken att denna typ av tester egentligen blir en mätning av elevernas förmåga när det gäller att överbrygga språk- eller kulturgränser och att resultaten från PISA vanligen inte kommuniceras utifrån en sådan slutsats.

Referenser

- Allerup, Peter (2007): Identification of group differences using PISA scales – considering effects of inhomogeneous items. I Stefan Hopmann; Gertrude Brinek & Michael Retzl. red.: *PISA According to PISA: Does PISA Keep what it Promises?* s 175–202. Wien: LIT
- Biesta, Gert (2010): *Good Education in an Age of Measurement: Ethics, Politics, Democracy*. Boulder: Paradigm Publishers.
- Grek, Sotiria (2009): Governing by numbers: the PISA ‘effect’ in Europe. *Journal of Education Policy* 24(1), 23–37.
- Fertig, Michael (2004): What can we learn from international student performance studies – Some methodological remarks. Essen: RWI Discussion Paper. No. 23..
- Jakobsson, Anders; Davidsson, Eva; Karlsson, Karl-Göran. & Oskarsson, Magnus (2013): Exploring epistemological trends in students’ understanding of science from the perspective of large-scale studies. *ISRN Education*. 13 (1), 4-18. (Open access: Article ID 196014).
- Mayer, Heinz-Dieter & Benavot, Aaron (2013): *PISA, power and policy – the emergence of global educational governance*. Oxford (UK): *Oxford Studies in Comparative Education*.
- OECD (2009): *PISA 2009 Assessment Framework - Key competencies in reading, mathematics and science*. Paris: OECD publications.
- SOU 2008:109 *En hållbar lärarutbildning*.