

Att pröva elevers läsförmåga i årskurs nio

En analys av uppgiftskonstruktioner i det nationella provet i svenska, delprov A *Att läsa och förstå*

Michael Tengberg

ASSESSING READING COMPREHENSION IN 9TH GRADE. AN ANALYSIS OF TEST ITEMS IN THE NATIONAL TESTS OF SWEDISH, PART A *TO READ AND TO COMPREHEND*. Over the last decade, national tests have become increasingly important to define educational quality. Testing procedures are more time consuming and take a prominent role as instruments for control and governance. This development necessitates critical analyses of prevalent test constructions and of test validity. The article reports from an analysis of test items and scoring guidelines in the national test of Swedish in 9th grade, focusing specifically on the measurement of literary comprehension. The analysis is based on tests used 2005–2012. Drawing on test theory and reading process theory, findings indicate that the construct definition of reading comprehension in these tests is deeply problematic and that alternative test constructions should be considered.

Keywords: assessment, construct validity, educational testing, item format, reading comprehension.

Inledning

Den här artikeln handlar om läsprov i skolan som ett instrument för bedömning av läsförmåga inom ramen för det nationella provet. Ett flertal studier har tidigare ägnats åt det muntliga delprovet (Brännström 2011, Hagberg-Persson 2009, Palmér 2009) och det

Michael Tengberg är lektor i pedagogiskt arbete vid Institutionen för pedagogiska studier, Karlstads universitet, 651 88 Karlstad. E-post: michael.tengberg@kau.se

skriftliga delprovet (Borgström 2012, Nyström Höög 2009, Östlund-Stjärnegårdh 2010), men ytterst lite forskningsintresse har ägnats åt läsprovet. Analysen i studien riktar sig mot provkonstruktionen som sådan och vill medels en explicit genomgång av texter, provuppgifter och bedömningsanvisningar för det första svara på vilken typ av uppgifter som kännetecknar proven och vilka slags läsprocesser dessa uppgifter ställer krav på. För det andra syftar analysen till att mot testteoretisk och lästeoretisk bakgrund blottlägga problem som dels har med provkonstruktionen som sådan att göra, dels med synen på text och läsning. Den fråga som väcks i artikeln är alltså den om läsprovets validitet i förhållande till kursplaner och vetenskapliga perspektiv. Som en avgränsning har jag valt att koncentrera analysen till läs-förståelseprovet i årskurs nio och till de provuppgifter som relaterar till de skönlitterära texterna i det till provet hörande texthäftet. De skönlitterära texterna, i synnerhet prosatexterna, utgör, som jag ska visa, en utrymmesmässigt betydande del av läsprovet, något som i sig motiverar ett närmare studium.

Bakgrund

Utbildningen i skolan ska vila på vetenskaplig grund och beprövad erfarenhet, heter det i Skollagens första kapitel. Detta krav rör dels urvalet av ämnesinnehåll, dels undervisningens uppläggning och organisering. Men kravet på vetenskaplig förankring torde förstås också gälla de kunskapsbedömningar i form av betyg och nationella prov, som under de senaste åren ökat såväl i omfång som i betydelse. Eva Forsberg och Christian Lundahl (2006) menar att kunskapsbedömningar kommit att få en alltmer framträdande roll som instrument för styrning och kontroll av skolan och kan därmed betraktas som "en aspekt av läroplanen" (s 8). I den utbildningspolitiska debatten duggar vidare referenserna till internationella jämförelser av elevers kunskaper tätt och blir snabbt till retoriska redskap för olika politiska viljeriktningar (jfr Pettersson 2008). På samma vis får offentliggörandet av kommuners och skolors resultat i de nationella proven betydelse för politiska och ekonomiska beslut liksom för strömningar inom det fria skolvalet. Kunskapsbedömningarna, menar Forsberg och Lundahl, antyder och styr hur olika kunskaper ska värderas (2006, s 23f). Detta gör det angeläget att kritiskt granska de kunskapsstyver och ämnessyner som legitimeras och premieras i såväl internationella kunskapsmätningar som inhemska nationella prov.

Ett tydligt exempel på hur internationella kunskapsmätningar bidragit till att sätta den utbildningspolitiska dagordningen i Sverige

finner man i de senaste årens omtalade kris när det gäller elevers läsförmåga. Sjunkande resultat för svenska elever i Programme for International Student Assessment (PISA) och Progress in International Reading Literacy Study (PIRLS) används för att motivera såväl statliga utredningar (SOU 2012:65) som satsningar på lokala läsprojekt och kompetensutvecklingstillfällen (se t ex Granström med flera 2012). De testresultat som refereras framstår lätt som alarmerande och ansvarskännande politiker tvingas följaktligen kavla upp ärmarna och visa handlingskraft, trots att kunskapen om problemets natur i själva verket är starkt begränsad. Huruvida svenska elever läser bra eller inte kan med referens till PISA-resultaten endast beläggas som ett *relativt* mått. Läs-krisen definieras av Sveriges gradvis försämrade placering på PISA:s rankinglistor, och inte på någon grundlig analys av skolans läsundervisning eller av vilket slags läsförmåga som liv och deltagande i det moderna samhället och i den moderna kulturen förutsätter (jfr Holm 2009, s 145). Sällan ifrågasätts i den allmänna skoldebatten om läsförmåga verkligen är ett fenomen som låter sig prövas bara lite sådär i allmänhet. Är det verkligen så att PISA:s definitioner av *reading literacy* – översatt av Skolverket som ”förmåga att förstå, använda, reflektera över och engagera sig i texter för att uppnå sina egna mål, utveckla sina kunskaper och sin potential och för att delta i samhället” (Skolverket 2010, s 29) – sätter samma aspekter av läsförmåga i fokus som kursplanerna och de nationella proven i svenska? I Skolverkets rapport från PISA-undersökningen 2009 fastslås att det finns en hög grad av samstämmighet mellan läsförmåga så som den definieras av kursplanemål och ämnesprov i svenska och läsförmåga så som den definieras av PISA (Skolverket 2010, s 35–37). Liknande slutsatser har dragits i andra Skolverksrapporter (se Skolverket 2006 och 2012a). Man bör förstås se dessa jämförelser som lovvärda initiativ, men de presenterade slutsatserna bygger varken på någon djuplodande analys av proven på item-nivå eller på någon rigorös sammanställning av provens sammansättning ifråga om läsprocesser, texttyper eller föreställd lässituation. De är alltså snarare att betrakta som inledande sonderingar av terrängen än som några tillfredsställande svar på den övergripande problemställningen om hur läsförmåga definieras och prövas i olika sammanhang.

Studiens syfte

Det övergripande syftet med den här studien är att bidra till en fördjupad medvetenhet om vilket slags läsförmåga i förhållande till skönlitterära texter som prövas i de nationella proven och därigenom

skapa möjlighet till en kritisk reflexion över läsproven som underlag för bedömning av elevers läsförmåga. Ett ytterligare syfte med studien är att bidra med utgångspunkter för relevanta jämförelser mellan bedömningsrationalen i de svenska läsproven och bedömningsrationaler i andra länders läsprov för motsvarande åldrar. Ett sådant underlag kan också ge bättre förutsättningar för att förstå likheter och skillnader mellan det slags mätning av läsförmåga som görs i de svenska nationella proven med dem som görs dels i PISA-testerna, dels i olika typer av diagnostiska material.

Forskningsfrågor

Följande forskningsfrågor har legat till grund för den empiriska analysen.

1. Hur ser fördelningen ut mellan andelen uppgifter som ges till skönlitterära texter respektive till andra texttyper inom ramen för det nationella provet, delprov A, *Att läsa och förstå*? Hur har detta förhållande förändrats över tid?
2. Vilka typer/kategorier av provuppgifter ges till de skönlitterära texterna och vilket slags läsförmåga prövar de?
3. Vilka principer råder i läsprovets bedömningsanvisningar, det vill säga vilka principer blir styrande för hur elevers responser på provuppgifterna ska bedömas?
4. Hur förhåller sig läsprovets uppgifter och bedömningsanvisningar till kursplanen i svenska? Mäter provet det som utbildningssystemet förutsätter/förväntar sig att det mäter?

Läsförståelseprov och testteoretiska parametrar

För tester av typen nationella prov är höga krav på reliabilitet och validitet självklara. Vi måste med andra ord kunna förlita oss på att de svar elever producerar utifrån provuppgifterna ger goda mått på de ämneskunskaper som kursplanerna i respektive ämne definierar. Att ta mått på läsförmåga är särskilt utmanande, eftersom läsförmåga sammanhänger med så många olika variabler (Snow 2002). Ordval, krav på elevers skriftliga formuleringsförmåga och differenser i elevers omvärldskunskaper är exempel på variabler som riskerar att hota reliabiliteten i läsförståelseprov (Francis et al 2006, Roe & Lie 2011). När det gäller de svenska nationella proven tillkommer det faktum att proven rättas av en mängd olika lärare, vilka sannolikt inte kommer att värdera elevers svar på samma sätt, trots att de tillämpar

samma bedömningsmallar. Detta faktum har varit föremål för en del debatt de senaste åren (se bland annat Skolinspektionen 2012). Andra reliabilitetsmått handlar om överensstämmelse mellan upprepade mätningar liksom överensstämmelse mellan frågor som är avsedda att mäta samma sak (Black & Wiliam 2012).

Tillförlitlighet i kunskapsmätningar betyder emellertid föga om vi inte först noggrant definierat den *kunskapsdomän* (Koretz 2008, s 19) som ska prövas (se även Wiliam 2010). Vad menar vi kort och gott med begreppet läsförmåga? För det nationella provets del handlar det således om att utgå från kursplanens definitioner av läsning och läsförmåga. Dessa är nu tämligen kortfattade och därmed svårtolkade. I de prov som gällde under Lpo94 gavs tillsammans med bedömningsanvisningarna ett tillägg i form av för läsförståelsedelen specifika betygsgränser. Dessa utgjorde alltså provkonstruktörens uttolkning av kursplanens definition av läsförmåga. I internationella mätningar som exempelvis PISA är ramverket och begreppsdefinitionen ofta mer explicit redogjorda för (se OECD 2009). Däremot har OECD hitintills varit påfallande restriktiva när det gäller att publicera testuppgifter och bedömningsanvisningar, vilket gör det svårt för utomstående att pröva och diskutera förhållandet mellan angiven kunskapsdomän och testinstrument. Eventuella brister i detta förhållande skulle innebära att provet eller testet lider av bristande *begreppsvaliditet*. Begreppsvaliditet ses här som en överordnad term, som också innesluter ovan nämnda aspekter av reliabilitet (jfr Black & Wiliam 2012). En utgångspunkt kan tas i Samuel Messicks definition: “[v]alidity is an integrative evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment.” (1989, s 13) Det betyder att validitetsprövning innefattar inte bara provkonstruktionens relation till den avsedda kunskapsdomänen utan även de tolkningar och konkreta användningar som görs av olika mätresultat. Validitet är med andra ord inte *en* egenskap som kan knytas till visst ett prov utan olika egenskaper beroende på i vilket syfte någon använder ett provresultat (Nuttall 1987, s 110).

Provuppgifter, textval och delskalor

Hur en uppgift formuleras liksom vilket responsformat som erbjuds är avgörande för vad ett prov egentligen mäter, och därigenom – med hänsyn till de inferenser som görs utifrån ett givet testresultat – avgörande för provets validitet. Åsa Wedin (2010) har analyserat lästestet ”Vilken bild är rätt?” för grundskolans tidigare år. I testet

får eleverna i var och en av uppgifterna läsa en mening och därefter bland fyra olika bilder välja den som bäst representerar innehållet i meningen. Wedin visar hur en måhända konventionell men knappast intuitivt given kunskap om bildsymbolik förutsätts för att eleven ska kunna nå fram till rätt svar. Motsägelser i sammansättningen av de fyra bilderna liksom ett potentiellt kulturellt främlingskap riskerar därmed leda till att testet mäter (och således begränsas av) sådant som inte hör till den valda domänen. Detta innebär ett hot mot validiteten (se Messick 1989, s 34f) och kan medföra att provet renderar felaktiga uppgifter om elevernas läsförståelse.

Att olika prov fångar olika aspekter av läsförmågan kan demonstreras genom att analysera det relativa bidraget till olika provresultat från sådana kognitiva processer som antas bygga upp fenomenet läsförståelse (ordavkodning, muntliga färdigheter, läshastighet och minne t ex). Laurie E. Cutting och Hollis S. Scarborough (2006) har visat hur sådana faktorer får olika betydelser i olika läsprov, vilket implicerar behovet av bättre insikter om vad olika tester egentligen mäter och i vilken mån de svarar mot avnämarnas förväntningar och tänkta inferenser. Likaså har Paul Deane med flera (2006) visat hur provtexters utformning definierar provets sätt att på ett specifikt vis rekonstruera fenomenet läsförståelse.

Ett generellt problem är att de flesta läsprov innehåller en rad olika underliggande variabler och därmed endast tar ett översiktligt mått på läsförståelsen, ett mått som inte nödvändigtvis ger lärare information om hur elevens läsutveckling ska stöttas på bästa sätt (Kintsch & Kintsch 2005). Två elever kan nå samma provresultat men på olika vägar. Prov som PISA består av flera delskalor, vilket i bästa fall kan bidra till mer detaljerad information om elevens kunskaper, men det är komplicerat att skapa uppgifter som inom ramen för respektive delskala korrelerar högre med varandra än med uppgifterna inom övriga delskalor (Roe & Lie 2011).

Uppgiftsformat

Det råder likväl en bred enighet om att läsprov bör täcka in flera olika processnivåer och texttyper. Mer debatt har det dock varit ifråga om responsformatet. I den psykometriska traditionen finns en utbredd acceptans för tester med *flervalsuppgifter* (*multiple choice*), det vill säga uppgifter med fasta svarsalternativ (i standardformatet ges fyra eller fem svarsalternativ av vilka ett är rätt och övriga felaktiga). Detta format brukar ställas mot öppna *uppgifter* (*constructed response*), det vill säga uppgifter som kräver att eleverna i skrift formulerar egna

svar och där läraren bedömer varje svar individuellt. I storskaliga tester finns det flera uppenbara fördelar med flervalssuppgifter: bedömningsproceduren är mycket mer reliabel och likvärdig; elevernas skrivfärdighet blandas inte in i bedömningen; utrymme ges för fler texter och fler frågor under loppet av ett prov, vilket gör att en större del av domänen kan täckas in av provet; sofistikerade statistiska analyser av typen Item Response Theory (IRT) kan användas för att skapa relevanta och reliabla delskalor av den övergripande domänen; flervalssuppgifter är mer kostnadseffektiva och ger en snabbare och enklare hantering av svarsmaterial och resultatrapportering (Campbell 2005; Olsen 2007; Roe & Lie 2011).

Samtidigt finns det kritik mot användningen av flervalssuppgifter i kunskapsbedömningar. Lauren B. Resnick och Daniel P. Resnick (1992) har exempelvis argumenterat för att användandet av flervalssuppgifter som testinstrument bygger på en förlegad kunskapssyn och att det testar kunskaper som en samling isolerade färdigheter eller faktakunskaper. En annan mycket vanlig invändning är att flervalssuppgifter förutsätter frågor med enskilda korrekta svar, vilket innebär att mer kvalificerade texttolkningsproblem inte kan prövas (Duke 2005; Langer 2011). Den invändningen måste emellertid göras även för öppna frågor när dessa kombineras med alltför slutna bedömningsanvisningar, vilket som vi ska se nedan i hög grad gäller det nationella läsprovet. Flervalssuppgifter kan också innehålla möjligheten till fler än ett korrekt svarsalternativ, eller utformas så att eleven ska gradera rimligheten i olika svarsalternativ, exempelvis på en skala 2–1–0 där 2 indikerar *mycket bra svar*, 1 indikerar *på rätt spår* medan 0 betyder *helt fel*. Empiriska prövningar har visat att de senare varianterna stimulerar till ett mer komplext kognitivt processande, som i högre grad korrelerar med den sortens läsprocesser som kan beläggas i öppna kvalitativa intervjuer. Sådana format möjliggör alltså prövande av en mer kvalificerad textförståelse (se Pearson & Hamm 2005, s 34–36). P. David Pearson med flera (1990) har också visat empiriskt att skillnaden mellan de kognitiva processer som stimuleras av flervalssuppgifter respektive av öppna uppgifter generellt är mycket liten. Några typer av öppna uppgifter, där elever ombads reflektera över, värdera och jämföra litterära texter (poesi), visade sig dock vara bättre på att stimulera dels mer nyansrika perspektiv, dels intertextuella perspektiv. Lärdomen av dessa studier är att om man verkligen vill pröva djup läsförståelse är det bättre att konstruera äkta öppna uppgifter istället för att göra om konventionella flervalssuppgifter till kortsvarsuppgifter (Pearson & Hamm 2005).

Ett flertal studier har också visat att flervalssuppgifter i högre grad gynnar pojkar medan flickors resultat gynnas av öppna uppgifter

(Breland 1991; Lafontaine & Monseur 2006; Lissitz & Hou 2012). Mot bakgrund av att elever ofta lämnar många uppgifter obesvarade, kanske till följd av upplevd oförmåga att lösa en viss uppgift, är det värt att notera att pojkar i högre utsträckning svarar på flervalsuppgifter än på öppna uppgifter. Det gäller även svaga elever (Olsen 2007). Det tycks alltså finnas en validitetsproblematik som relaterar till responsformatet.

Metod och material

Föreliggande studie bygger dels på en enkel statistisk sammanställning av förekommande texter, texttyper och provuppgifter i läsprovet, dels på en innehållsanalys av provuppgifter och bedömningsanvisningar utifrån sex olika variabler.

Urval

Det studerade materialet utgörs av delprov A *Att läsa och förstå* åren 2005–2012. Från och med 2013 har hela det nationella provet förändrats eftersom detta års niondeklassare är de första som läser utifrån Lgr11. Av det skälet har jag valt att inte ta med provet från 2013. Att försöka ge någon övergripande bild av det nuvarande provets utformning och innehåll är dessutom naturligtvis alldeles för tidigt. Det material som studeras har alltså använts under Lpo94. Eftersom det fortfarande är samma arbetsgrupp som utformar proven och eftersom stora delar av provsystemet förblivit sig likt, menar jag att en analys av de prov som använts under det gångna decenniet fortfarande är av såväl akademisk som allmän relevans.

Variabler

Varje provuppgift med tillhörande bedömningsanvisning har kategoriserats utifrån de sex variablerna i syfte att ge en allsidig bild av uppgifternas och bedömningens innehållsliga fokus, processorientering och tekniska utformning. Därtill har jag undersökt i vad mån eleverna informeras om bedömningens inriktning. De undersökta variablerna fokuserar *uppgiftsformat*, *läsprocessnivå*, *innehållsfokus*, *svarsutrymme*, *information till eleverna* samt *bedömningsnivåernas utgångspunkter*. Uppgiftsformat har diskuterats utförligt ovan. Här skiljer jag på flervalsuppgifter, öppna uppgifter och kombinationsfrågor som innehåller båda formaten.

Läsprocess och innehållsfokus

För att analysera ämnessyn och kunskapssyn i ett bedömningsinstrument är det rimligt att utgå från en i sammanhanget relevant kunskapstaxonomi i syfte att peka ut uppgifters kravnivåer och innehållsfokus utifrån en etablerad kunskapsteori (jfr Wedin 2010; Näsström 2008). En del av den här analysen utgår därför från den uppdelning i tre övergripande läsprocesskategorier som OECD använder vid utformningen av PISA-testet. De här tre aspekterna ”are the *mental strategies, approaches or purposes that readers use to negotiate their way into, around and between texts.*” (OECD 2009, s 34) De benämns på följande vis: *access and retrieve* (att finna, urskilja och använda rätt information i texten); *integrate and interpret* (att föra ihop olika element i texten, tolka och dra slutsatser om sådant som inte är explicit uttalat); och *reflect and evaluate* (att knyta aspekter av texten till egna erfarenheter, kunskaper och idéer för att göra jämförelser och för att värdera innehåll och form i texten) (s 35–39).¹ Ett uppenbart skäl för att använda just PISA:s lästeoretiska ramverk för att kategorisera uppgiftstyperna i nationella provet är att det ger möjlighet till en initial jämförelse av läsprocessfokus i de två proven. Ett annat skäl är att PISA:s ramverk bygger på ledande internationell forskning om läsförståelse och används i såväl läsförståelsestudier som i andra länders kurs- eller läroplaner.

I analysen av uppgifternas innehållsfokus har jag däremot gått induktivt tillväga och formulerat kategorierna utifrån det befintliga materialet. Den analysen ska ses som ett komplement till de kategoriseringar som görs med avseende på läsprocessnivå.

Bedömningsrational

Resterande tre variabler avser delar av bedömningens bakomliggande rational, det vill säga på vilka logiska eller teoretiska grunder bedömningsanvisningen till den rättande läraren tycks vila. Intresset för svarsutrymmet handlar om den grad av frihet som ges inom bedömningsanvisningen för att bedöma ett svar som korrekt eller icke korrekt. När det gäller information till eleverna undersöks huruvida eleverna i provhäftet informeras om på vilka grunder deras svar kommer att bedömas. Avslutningsvis undersöks om bedömningen av de högre nivåerna (2–3)² bygger på en kvalitativ eller kvantitativ bedömningssyn. Jag har i en annan studie (Tengberg under utgivning) visat hur uppgifter som förekommer i det nationella läsprovet bygger på en kvantitativ bedömningssyn. Detta genom att bedömningsanvisningen premierar att eleven fått med en viss *mängd*

aspekter av ett förväntat svar, snarare än att svaret ska återspegla en viss kvalitet i läsningen/tolkningen av texten. Här undersöks alltså detta fenomenets utbredning i de förekommande läsproven.

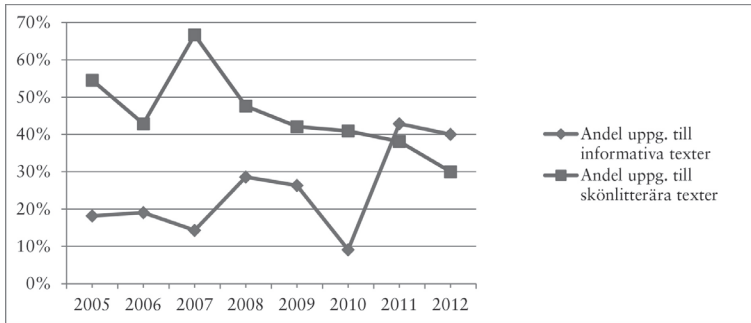
Resultat

Skönlitteraturen i det nationella provet

De texter som förekommer i proven 2005–2012 är inte kategoriserade utifrån texttyp. Utifrån traditionella genreavgränsningar går det emellertid att avgränsa gruppen skönlitterära texter till i huvudsak noveller, romanutdrag och dikter. Övriga texttyper kan kategoriseras som informativa texter/nyhetstexter, krönikor, recensioner, insändare/debatterande texter, brev samt tabeller och diagram.³ I analysen av hur andelen texter och uppgifter fördelas mellan olika texttyper har jag endast räknat de texter till vilka det ingår provuppgifter. I varje års prov finns ofta också en eller ett par uppgifter av typen ”para ihop en bild och en text under rubriken /.../ motivera varför dessa passar ihop”. Dessa uppgifter innebär att alla provhäftets texter *kan* (beroende på elevens val) förekomma i uppgiften, men för att ge en rättvisande bild av hur fördelningen av antalet uppgifter på respektive texttyp ser ut har jag utelämnat dessa uppgifter vid beräkningen. Observera att noveller och romanutdrag alltid är de i särklass längsta texterna i texthäftet, vilket innebär att dessa texter med vidhängande provuppgifter upptar en proportionellt större del av provandet av elevers läsförmåga i det nationella provet än vad *antalet* texter avslöjar. En bättre indikator blir därför antalet provuppgifter.

Totalt under undersökningsperioden upptar de skönlitterära texterna med tillhörande provuppgifter en betydande men också gradvis minskande andel av det totala antalet texter och uppgifter (se figur 1). Som diagrammet visar har istället de informativa texterna med tillhörande uppgifter kommit att få större plats i läsprovet. Förändringen i detta förhållande är tydlig och särskilt intressant med tanke på att samma läroplan och kursplan hela tiden legat till grund för provutformningen.

Figur 1. Andel uppgifter till skönlitterära respektive informativa texter i läsprovet



Inom gruppen uppgifter till skönlitterära texter dominerar andelen uppgifter knutna till prosatexter helt över andelen uppgifter knutna till dikter. Endast i två av proven ställs en direkt fråga till en enskild dikt. Däremot förekommer i alla utom ett av proven uppgifter som innebär att eleven ska para ihop antingen två av texterna eller en bild och en text i häftet (ibland anges att texten ska vara en dikt) och förklara varför hen tycker att dessa texter/bilder hör ihop. Uppgiften kan också formuleras så att en text ska kombineras med ett ordspråk eller med texthäftets titel och förklara varför texten utgör en bra representation av ordspråket/titeln.

Uppgiftstyper och innehållsfokus

Fördelningen av olika uppgiftstyper i det nationella läsprovet är inte explicit utarbetad efter något teoretiskt ramverk över olika läsprocesser. Som angavs ovan presenteras dock några kriterier (fritt tolkade från kursplanekriterierna) för de olika betygsnivåerna. Dessa varierar något över undersökningsperioden, men sedan 2007 anges i lärarinformationen att provet prövar förmågan att ”tillämpa olika lässtrategier (sök läsning, översiktsläsning, jämförande och tolkande läsning) och uttrycka sina egna reflektioner och åsikter kring texter.” (Skolverket 2011, s 8). Dessa kategorier kan grovt relateras till PISA:s läsprocessnivåer, även om de senare erbjuder en delvis annan bild. Som framgår av tabell 1 är sök läsning och tolkande läsning tydligt representerande bland uppgifterna till skönlitterära texter i det svenska nationella provet, medan endast ett fåtal uppgifter prövar elevernas förmåga att reflektera och utvärdera. Jämfört med fördelningen av uppgifter för att pröva olika läsprocesser i PISA-testet så innehåller det svenska nationella provet således en förskjutning mot att pröva

enklare sökläsningförmågor framför att pröva mer komplexa reflexionsuppgifter.

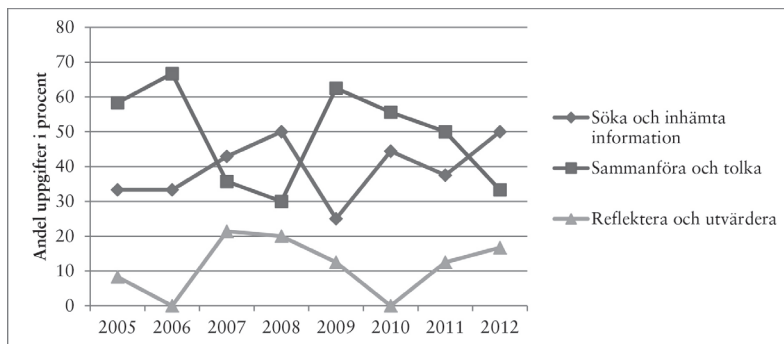
Tabell 1. Vilken läsprocess efterfrågas i uppgifterna?

Läsprocessnivå	Andel av uppgifter i NP 2005–2012	Andel av uppgifter i PISA 2000–2009
Söka och inhämta information	39,5	25
Sammanföra och tolka	48,7	50
Reflektera och utvärdera	11,8	25
Totalt	100,0	100

Källa rörande PISA 2009: OECD 2009 s 43

Studerar man hur fördelningen av uppgifter med avseende på läsprocessnivå sett ut över tid visar det sig att den varierat påtagligt (se figur 2). De dimensioner av läsförmåga som prövas ett år är med andra ord inte samma förmågor som prövas ett annat år. När fördelningen av uppgifter inom provet med avseende på både texttyp och avsedd läsprocessnivå varierar kraftigt är det tvärtom så att provet mäter olika slags kunskaper från till år.

Figur 2. Andel uppgifter på respektive läsprocessnivå 2005–2012



Det är emellertid inte möjligt att hävda att läsprocessnivåer beskrivna på det här viset skulle representera separata, mentala strategier hos eleverna, eller för den delen att varje enskild uppgift kan förutsättas involvera endast ett slags läsprocess (jfr Roe & Lie 2011; van Steensel med flera 2012). Tvärtom förutsätter ju reflexion och utvärdering någon föregående tolkning av texten, vilket i sin tur bygger på sammanföring av textens olika delar och som grund för detta även en läsning som söker

och lokaliserar viss information. På omvänt vis förutsätter sökläsandet ett visst inslag av tolkning och urskiljande. Ändå kan dessa kategorier ge en grov bild av vari provuppgifternas huvudsakliga utmaning ligger med avseende på läsprocess.

Betoningen på att söka information eller fakta i texterna bekräftas emellertid i analysen av uppgifternas innehållsliga fokus, det vill säga till vilka dimensioner i texterna som uppgifterna tycks relatera. Av tabell 2 nedan framgår att just uppgiften att finna enkla fakta i texterna upptar en stor del av provet. Även om några av dessa uppgifter ytligt sett tycks pröva förståelse av en situation eller en karaktärs handlingar, är situationer och handlingar redan tolkade; elevens uppgift blir istället att hitta stöd för att denna tolkning är riktig. Här ges två exempel på sådana uppgifter, dels till ett romanutdrag ur *Inga bomber över Skärholmen* (prov 2005), dels till novellen "Liten julnovell" (prov 2006).

När Rashid får ett telefonsamtal märks det att han inte vill att Peter ska lyssna eller förstå samtalet. Hur märks det? Ge något exempel (ur 2005 års prov).

Bettans julklapp till Benny (ett mål mat och ett hem för en natt) är planerad långt i förväg. Hur förstår man det av texten (ur 2006 års prov)?

Uppgiften består på det här viset snarare i att *finna textfakta* än i att *jämföra karaktärer* eller *tolka situationsbeskrivningar*. Det är med andra ord inte elevens förmåga att dra slutsatser och göra tolkningar av de skönlitterära texterna som prövas utan elevens förmåga att motivera/finna stöd för provkonstruktörens tolkning samt att i skrift återge detta stöd/motiv.

Tabell 2. Vilket slags innehåll riktar sig provuppgiften mot?

Innehållsfokus	n	%
Finna textfakta	26	34,2
Ge karaktärsdefinition/jämföra karaktärer med varandra	10	13,2
Använda omvärldskunskaper	2	2,6
Identifiera karaktärens handlingar/motiv	11	14,5
Urskilja tidsförlopp/berättarstruktur	4	5,3
Tolka situationsbeskrivning	13	17,1
Använda ord- eller språkkunskap	5	6,6
Ge texten en övergripande tolkning	5	6,6
Totalt	76	100,0

Proven innehåller vidare tre olika typer av uppgiftsformat: flervalsuppgifter, öppna frågor samt kombinationer av dessa två. Fördelningen mellan de olika uppgiftsformaten är intressant. I forskningslitteraturen lyder som vi sett ovan det huvudsakliga argumentet att öppna uppgifter är problematiska av flera skäl, men värdefulla i syfte att ge eleven chans att explicitgöra sina resonemang, sin förmåga att reflektera, granska, jämföra och kritiskt förhålla sig till texter (Campbell 2005; Roe & Lie 2011).

Tabell 3. Fördelning av uppgiftsformat

Uppgiftsformat	n	%
Flervalsfrågor	14	18,4
Öppna frågor	57	75,0
Kombination flervals- + öppna	5	6,6
Totalt	76	100,0

Som framgår av tabell 3 är en mycket hög andel av frågorna i nationella provet formulerade som öppna uppgifter där eleven förväntas producera antingen kortare eller längre skriftliga svar. Som jämförelse kan noteras att PISA 2009 endast till 43% består av uppgifter som kräver "expert judgement in coding", det vill säga öppna uppgifter (OECD 2009, s 46) samt att det norska ramverket för nationella prov sätter gränsen för andelen öppna uppgifter vid 35% (Utdanningsdirektoratet 2010, s 12). Här utmärker sig alltså den svenska provkonstruktionen. Man använder en jämförelsevis

mycket hög andel öppna frågor, detta trots att en, som vi såg ovan, stor andel av uppgifterna begränsar sig till att pröva elevernas förmåga att lokalisera och återge explicit information från texterna. Det öppna uppgiftsformatet används med andra ord inte huvudsakligen för att ge eleverna utrymme för att resonera, reflektera, diskutera och jämföra. En enkel korstabulering visar att 77% av de uppgifter som prövar förmågan att söka och inhämta information och 68% av uppgifterna som prövar förmågan att sammanföra och tolka utformats som öppna frågor (se tabell 4).

Tabell 4. Andel av olika uppgiftsformat fördelat på läsprocessnivå

Läsprocessnivå	Uppgiftsformat	Andel flervalsfrågor	Andel öppna frågor	Andel flerv. + öppna
Söka och inhämta information		23,3	76,7	0
Sammanföra och tolka		18,9	67,6	13,5
Reflektera och utvärdera		0	100	0
Totalt		18,4	75,0	6,6

De öppna uppgifterna tycks alltså närmast utgöra en standardkonstruktion i det svenska läsprovet. Med tanke på den debatt som förts de senaste åren om brister i likvärdighet mellan lärares bedömningar av de nationella proven finns det anledning att ifrågasätta om denna fördelning mellan olika uppgiftsformat är lämplig.

Bedömningsrational

En vidhängande fråga rör i vad mån uppgifters innehållsliga inriktning och format relaterar till eller motiverar vissa bedömningsrationaler. En av de undersökta variablerna gäller med vilken specificitet acceptabla svar avgränsas i provens bedömningsanvisningar. Som grund för analysen har jag urskilt tre olika frihetsgrader: anvisningar som anger att endast ett bestämt svar godkänns; anvisningar som godkänner några bestämda svarsalternativ samt anvisningar om individuell bedömning inom givna ramar. Den första kategorin är som framgår av tabell 5 nedan klart överrepresenterad.

Eftersom proven, som det framgått ovan, i stor utsträckning prövar förmågan att lokalisera vad som antas vara direkt utläsbar information ur texten kan det framstå som logiskt att en hög andel av svaren har direkt facitliknande svar. Å andra sidan kan det framstå som problematiskt att elevers läsning av just skönlitterära texter över

huvud taget i så hög utsträckning prövas som om det vore främst en informationssökande aktivitet där givna svar verkligen existerar. Som också framgår i tabell 5 anger bedömningsanvisningarna regelmässigt bestämda svar även på frågor på Sammanföra och tolka-nivån; detta samtidigt som en hög andel av dem alltså utgör öppna frågor på vilka elever formulerar längre skriftliga svar.

Tabell 5. Bedömningsanvisningarnas tolkningsutrymme fördelat på läsprocessnivå

Bedömningsanv. Läsprocessnivå	Bestämt svar anges	Några bestämda alternativ anges	Individuell bedömning inom givna ramar
Söka och inhämta information	76,7	13,3	10,0
Sammanföra och tolka	81,1	18,9	0
Reflektera och utvärdera	22,2	11,1	66,7
Totalt	72,4	15,8	11,8

Denna rationalitetskonflikt är särskilt besvärande eftersom gängse litteraturvetenskapliga och litteraturdidaktiska textteorier sedan lång tid tillbaka tämligen samfällt menar att när det gäller skönlitteratur existerar inte några 'korrekta' tolkningar eller några tolkningsperspektiv som är objektivt tillgängliga för alla läsare (eller ens för alla kompetenta läsare) (se Attridge 2004; Malmgren 1983; Palm 2002). I en kritisk uppgörelse med några av provuppgifterna i 2008 års prov argumenterade Mikael Nordenfors (2008) för att bedömningsanvisningarna till proven inte nödvändigtvis byggde på de enda möjliga läsningarna av de skönlitterära texterna. "Med öppna frågor bör logiskt sett en öppen syn även på nivåer och bedömning följa", menade Nordenfors med hänvisning till frågor på tolkningsnivå där endast ett bestämt svar accepteras (s 75). Jag har själv i en annan studie (Tengberg under utgivning) visat på hur provkonstruktörens ibland väl snäva tolkningar och långsökta inferenser skapar både praktiska problem i den enskilda bedömningen och validitetsproblem i förhållande till andra bedömningar i skolan (exempelvis betygsättningen). Dessa problem i skärningspunkten mellan textteori och testteori aktualiserar frågan om vilka slags syften provet egentligen ska uppfylla: likvärdighet i bedömning eller ämnesdidaktisk utveckling genom att konkretisera kursplanernas mål och betygskriterier.

Nästa fråga gällde i vad mån elever informeras om på vilka kriterier deras svar bedöms. Framgår bedömningskriterierna i uppgiften,

eller är det först efteråt som eleverna informeras om vilket slags svar provkonstruktören var ute efter? Ett exempel är uppgifter som uppmanar eleverna att motivera en tolkning med stöd i texten, där bedömningsanvisningen anger för den rättande läraren att två eller flera olika motiveringar ska lämnas för att ett elevsvar ska erhålla en viss bedömningsnivå. Detta utan att eleverna i provsituationen informeras om dessa bedömningsvillkor. I de undersökta proven är det fler än hälften (54%) av uppgifterna till skönlitterära texter som på liknande sätt döljer en del av bedömningsgrunden för eleverna.

Ett annat exempel utgörs av de tvådelade uppgifter på vilka rätt svar på a gör att svaret som helhet bedöms ligga på nivå 1, rätt svar på b gör att svaret ligger på nivå 2 och rätt svar på båda deluppgifterna ger bedömningsnivå 3. Denna uppgiftstyp, som sällan informerar eleven om bedömningsgrunderna, är också problematisk eftersom nivå 3-bedömningen därmed inte bygger på att eleven i kvalitativ mening tillfört något utöver det som krävs på nivå 1 och 2. Istället är det summeringen av två korrekta svar på a och b som genererar nivå 3: $1+2=3$. Eftersom de tre bedömningsnivåerna blir avgörande för elevernas möjligheter att (enligt det dåvarande betygssystemet) nå betygen Godkänt, Väl godkänt respektive Mycket väl godkänt (se textanvisningar för helhetsbedömning av delprovet i Skolverket 2011b, s 4) kan man kanske säga att $G+VG=MVG$! Detta harmonierar visserligen med den kvantitativa syn som präglar instruktionerna för hur resultaten på delproven ska sammanvägas till ett enda provbetyg (s 82), men det rimmar inte särskilt väl med den kunskapssyn som låg till grund för Lpo 94, i vilken kunskapsbedömningar *inte* skulle vara kvantitativt orienterade (Skolverket 2005, s 18) utan där högre betygsnivåer skulle representera ett mer kvalificerat kunnande (s 24).

Vad som menas med ett mer kvalificerat kunnande kan förvisso diskuteras, och kvantitet kan möjligen ses som ett slags kvalitet bland andra kvaliteter. Om man granskar de uppgifter som medger högre bedömning än nivå 1 (38 st) kan man dock se att det i merparten av fallen (22/38) är just kvantitativt utvecklade svar som ger högre bedömning. Med kvantitativt utvecklade svar menas att eleven fått med fler aspekter av en viss fråga, funnit två till skillnad från ett stöd i texten för en viss tolkning eller svarar rätt på ett visst antal frågor i en flerdelad uppgift. Bedömningsanvisningarna premierar alltså i högre grad mer omfattande eller mer fullständiga svar än svar av högre kvalitet.

Diskussion

När politiker, medier och verksamma inom utbildningssystemet tävlar om att beklaga sig över en hel generations bristande läskunnighet finns det anledning att granska närmare hur denna läskunnighet egentligen prövas inom ramen för det egna nationella provsystemet. Skolinspektionen (2012) har återkommande rapporterat att prov som lämnar stort utrymme för bedömarnas tolkningar medför reliabilitetsproblem. Den här undersökningen visar att läsförståelseproven också riskerar flera olika typer av validitetsproblem.

Om texttyper och uppgiftstyper varierar påtagligt över åren betyder det mot den testteoretiska och lästeoretiska bakgrund som diskuterats ovan att provets representation av läsförståelse ändras från år till år. Detta är anmärkningsvärt eftersom samtliga prov i undersökningsperioden genomförts utifrån en och samma kursplan och läroplan. Vilka motiv finns då för att successivt variera begreppsdefinitionen av det kunskapsområde som provet är satt att pröva? Den tidigare forskningen om utformning av läsprov understryker vidare att den jämförelsevis höga proportion av öppna frågor som det svenska nationella läsprovet består av riskerar, utöver minskade möjligheter till reliabla bedömningar, att medföra *irrelevant testvarians* (Messick 1989) genom att formatet dels gynnar somliga grupper av läsare mer än andra, dels riskerar att göra den skriftliga formuleringsförmågan till en begränsande faktor för vilken läsförmåga som kan uppvisas. Eftersom öppna frågor dessutom är mer tidskrävande måste provet innehålla färre texter och färre uppgifter, vilket kan innebära att det blir svårare att täcka av den avsedda kunskapsdomänen. Då har vi att göra med en *underrepresentation* (Messick 1989) av domänen, vilket också utgör ett konkret hot mot validiteten. Studier har också visat att öppna frågor som regel inte ger bättre förutsättningar än flervalsfrågor för att fånga läsförmågor av högre komplexitet (Campbell 2005; Kastner & Stangla 2011), i synnerhet inte när de i hög utsträckning används för att pröva förmågan att hitta fakta i texter och endast i liten utsträckning prövar förmågan att jämföra, reflektera och utvärdera.

I läsprovet som genomfördes 2013 och som bygger på den nya kursplanen har provkonstruktörerna närmast sig det teoretiska ramverket för läsförmåga i PIRLS och definierar de läsprocesser som prövas i nära överensstämmelse med hur de definieras i PIRLS ramverk. Ett problem med detta förfarande är dock att det saknar stöd i den nya kursplanen. Till skillnad från exempelvis i Norge, där såväl läroplan som läsprov sedan 2006 uttryckligen tar sin utgångspunkt i PISA:s ramverk, har man i kursplanen för svenska inte definierat mål

för läsning och läsförmåga i enlighet med PIRLS ramverk. Man kan givetvis invända att det inte ligger någon självklarhet i att anpassa nationella kursplaner till de sätt att definiera eftersträvarvärda kunskaper som används i internationella jämförande mätningar.⁴ Det vore rentav direkt förkastligt om skälet till en sådan anpassning vore att förbättra landets möjligheter till bra resultat på PIRLS- eller PISA-proven. I det norska fallet konstaterar man emellertid att grunden för deras anpassning snarare består dels i att man gjort bedömningen att PISA:s ramverk baseras på solid internationell expertis på fältet och därmed framstår som relevant, dels i att det i Norge inte tidigare existerat något motsvarande ramverk för läsförmåga (Roe & Lie 2011, s 152). I Sverige görs alltså denna anpassning endast som en del av provsystemet, men saknar grund i kursplanen, något som förstås riskerar leda till allvarliga problem med provets validitet.

I takt med att kostnaderna i tid och pengar utövar ett allt hårdare tryck på det system som ska genomföra kunskapsproven i skolan vore det angeläget att hitta lösningar som är mer kostnadseffektiva än de vi använder idag. Ett övergripande syfte med de nationella proven i Sverige är att bidra till en likvärdig bedömning av elevers ämneskunskaper. Det får dock inte betyda att kunskapsdomänen reduceras till sådant som är enkelt att mäta med hög reliabilitet. Det finns trots allt goda argument för att använda uppgifter som låter elever demonstrera sitt kreativa och kritiska tänkande, som låter dem visa att de kan växla mellan olika perspektiv och jämföra olika texter (jfr Campbell 2005). Men som framgått ovan finns det också goda skäl att fundera över alternativ till den rådande provkonstruktionen.

Noter

1. I den svenska PISA-rapporten översätts läsprocessdimensionerna på följande vis: *access and retrieve* (söka och inhämta information); *integrate and interpret* (sammanföra och göra en tolkning); och *reflect and evaluate* (reflektera och utvärdera) (Skolverket 2010, s 34). Dessa översättningar används i den här artikeln.
2. Skolverket (2012b) ger följande definition: ”I delprovet finns uppgifter som kan ge svar på tre olika nivåer. Nivå 1 representeras vanligen av kortsvar som kan utläsas direkt ur texten. Nivå 2 ger svar som kräver att eleven kan överblicka flera avsnitt i texten eller visar visst mått av abstrakt tänkande, och nivå 3 kräver utförligare svar där flera aspekter på uppgiften efterfrågas.” (s 45).
3. Tabeller och diagram förekommer oftast tillsammans med en informativ text/nyhetstext.
4. En kritik mot användandet av PIRLS och PISA:s ramverk för läsförmåga kan vara att de tar liten hänsyn till att det är just skönlitteratur som läses. Istället tycks båda dessa ramverk, åtminstone med avseende på just läsprocess, förutsätta att samma läsprocesser, det vill säga samma mentala strategier, är i bruk vid läsning av skönlitteratur som vid läsning av informativa texter. Detta motsägs av internationell empirisk läsprocessforskning (Bortolussi & Dixon 2003, Miall 2006).

Referenser

- Attridge, Derek (2004): *The Singularity of Literature*. Oxon: Routledge.
- Black, Paul & Wiliam, Dylan (2012): The reliability of assessments. I John Gardner, red: *Assessment and Learning* (2:a uppl.), s 243–263. London: SAGE.
- Bortolussi, Marisa & Dixon, Peter (2003): *Psychonarratology: Foundations For the Empirical Study of Literary Response*. Cambridge: Cambridge University Press.
- Borgström, Eric (2012): Skrivförmåga på prov. I Gustaf Skar & Michael Tengberg, red: *Svenskämnet i går, i dag, i morgon*, s 209–223. Stockholm: Svenskläraryöreningen.
- Breland, Hunter M. (1991): *A Study of Gender and Performance on Advanced Placement History Examinations* (College Board Report No. 91-4, ETS Research Report No. 91-16). New York: Entrance Examination Board.
- Brännström, Maria (2011). *Bedömning av det muntliga nationella provet: En mötesplats för metaspråk om muntligheten*. Malmö: Malmö högskola.

- Campbell, Jay R. (2005): Single instrument, multiple measures: Considering the use of multiple item formats to assess reading comprehension. I Scott G. Paris & Steven A. Stahl, red: *Children's Reading Comprehension and Assessment*, s 347–368. Mahwah, New Jersey: Lawrence Erlbaum Ass.
- Cutting, Laurie E. & Scarborough, Hollis S. (2006): Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading* 10(3), 277–299.
- Deane, Paul; Sheehan, Kathleen M.; Sabatini, John; Futagi, Yoko & Kostin, Irene (2006): Differences in text structure and its implications for assessment of struggling readers. *Scientific Studies of Reading* 10(3), 257–275.
- Duke, Nell K. (2005): Comprehension of what for what: Comprehension as a nonunitary construct. I Scott G. Paris & Steven A. Stahl, red: *Children's Reading Comprehension and Assessment*, s 93–108. Mahwah, New Jersey: Lawrence Erlbaum Ass.
- Forsberg, Eva & Lundahl, Christian (2006): Kunskapsbedömningar som styrmedia. *Utbildning & Demokrati – tidskrift för didaktik och utbildningspolitik* 15(3), 7–29.
- Francis, David J.; Snow, Catherine E.; August, Diane; Carlson, Coleen D.; Miller, Jon & Iglesias, Aquiles (2006): Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading* 10(3), 301–322.
- Granström, Malin m fl (2012): Ett språklyft är nödvändigt. I NWT 2012-12-17.
- Hagberg-Persson, Barbro (2009): Elevers muntliga språkbehärskning i tidiga skolår. Hur kan muntliga färdigheter observeras, diagnostiseras och prövas? I Anne Palmér, red: *Muntlighetens möjligheter – retorik, berättande, samtal. Sjätte nationella konferensen i svenska med didaktisk inriktning, Uppsala 27–28 november 2008*, s 37–46. Uppsala: Institutionen för nordiska språk.
- Holm, Lars (2009): Who evaluates the evaluation? An analysis of computer-adaptive testing using literacy as an example, *Nordisk pedagogik* 29(1), 137–148.
- Kastner, Margit & Stangla, Barbara (2011): Multiple choice and constructed response tests: Do test format and scoring matter? *Procedia – Social and Behavioral Sciences* 12, 263–273.

- Kintsch, Walter & Kintsch, Eileen (2005): Comprehension. I Scott G. Paris & Steven A. Stahl, red: *Children's Reading Comprehension and Assessment*, s 71–92. Mahwah, New Jersey: Lawrence Erlbaum Ass.
- Koretz, Daniel (2008): *Measuring Up: What Educational Testing Really Tells Us*. Cambridge MA: Harvard U.P.
- Lafontaine, Dominique & Monseur, Christian (2006): *Impact of Test Characteristics on Gender Equity Indicators in the Assessment of Reading Comprehension*. University of Liège: Liège.
- Langer, Judith A. (2011). *Envisioning Literature: Literary Understanding and Literature Instruction (2:a uppl.)*. New York: Teachers College Press.
- Lissitz, Robert W. & Hou, Xiaodong (2012): The contribution of constructed response items to Large Scale Assessment: Measuring and understanding their impact. *Journal of Applied Testing Technology* 13(3), 1–50. <http://www.testpublishers.org/assets/volume%2013%20issue3%20constructed%20response.pdf> Hämtad 2013-05-27.
- Malmgren, Lars-Göran (1983): *Litteraturreception och litteraturpedagogik: Utgångspunkter för ett forskningsprojekt*. Lund: LU.
- Messick, Samuel (1989): Validity. I Robert L. Linn, red: *Educational Measurement (3:e uppl.)*, s 13–103. New York: Macmillan.
- Miall, David S. (2006): *Literary Reading. Empirical and Theoretical Studies*. New York: Peter Lang Publishing.
- Nordenfors, Mikael (2008): Ord är vaga och mångtydiga. *Pedagogiska magasinet nr 4/08*, 74–75.
- Nuttall, Desmond L. (1987): The validity of assessments. *European Journal of Psychology of Education* 2(2), 109–118.
- Nyström Höög, Catharina (2010): *Mot ökad diskursivitet? Skrivutveckling speglad i provtexter från årskurs 5 och årskurs 9*. Uppsala: Institutionen för nordiska språk.
- Näsström, Gunilla (2008): *Measurement of Alignment between Standards and Assessment*. Diss. Umeå: Umeå universitet.
- OECD (2009): *PISA 2009. Assessment Framework: Key competencies in reading, mathematics and science*. <http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/pisa2009assessmentframework-keycompetenciesinreadingmathematicsandscience.htm> Hämtad 2013-05-23.
- Olsen, Rolf V. (2007): Flervalsoppgaver: En papegøye fra Amerika? *Utdanning* 16, 44–45.

- Palm, Anders (2002): Att tolka texten. I Staffan Bergsten, red: *Litteraturvetenskap: En inledning*, s 189–204. Lund: Studentlitteratur.
- Palmér, Anne (2009): Rättvis bedömning? Bedömaröverensstämmelse vid muntliga nationella prov i gymnasieskolan. I Anne Palmér, red: *Muntlighetens möjligheter – retorik, berättande, samtal. Sjätte nationella konferensen i svenska med didaktisk inriktning, Uppsala 27–28 november 2008*, s 119–129. Uppsala: Institutionen för nordiska språk.
- Pearson, P. David, Greer, E.A., Commeyras, M., Stallman, A., Valencia, S.W., Krug, S.E., med flera (1990): *The Validation of Large Scale Reading Assessment. Building Tests for the Twenty-First Century*. Urbana, IL: University of Illinois, Center for the Study of Reading.
- Pearson, P. David & Hamm, Diane N. (2005): The assessment of reading comprehension: A review of practices: Past, present, and future. I Scott G. Paris & Steven A. Stahl, red: *Children's Reading Comprehension and Assessment*, s 13–70. Mahwah, New Jersey: Lawrence Erlbaum Ass.
- Pettersson, Daniel (2008): *Internationell kunskapsbedömning som inslag i nationell styrning av skolan*. Diss. Uppsala: Uppsala universitet.
- Resnick, Lauren B. & Resnick Daniel P. (1992): Assessing the thinking curriculum: New tools for educational reform. I Bernard R. Gifford & Mary Catherine O'Connor, red: *Changing Assessment: Alternative views of Aptitude, Achievement and Instruction*, s 37–76. Boston: Kluwer Academic.
- Roe, Astrid & Lie, Svein (2011). Nasjonale leseprøver i et didaktisk og testteoretisk perspektiv. I Stephen Dobson, Astrid B. Eggen & Kari Smith, red: *Vurdering, prinsipper og praksis: Nye perspektiver på elev- og læringsvurdering*, s 145–165. Oslo: Gyldendal.
- Skolinspektionen (2012): *Lika för alla? Omrättning av nationella prov i grundskolan och gymnasieskolan under tre år*. Stockholm: Skolinspektionen.
- Skolverket (2005): *Bedömning och betygsättning. Kommentarer med frågor och svar*. Stockholm: Fritzes.
- Skolverket (2006): *Med fokus på läsförståelse: en analys av skillnader och likheter mellan internationella jämförande studier och nationella kursplaner*. Stockholm: Skolverket.
- Skolverket (2010): *Rustad att möta framtiden? PISA 2009 om 15-åringars läsförståelse och kunskaper i matematik och naturvetenskap*. Rapport 352. Stockholm: Fritzes.

- Skolverket (2011a): *Ämnesprov Svenska/Svenska som andraspråk 2011. Delprov A. Att läsa och förstå*. Stockholm: Skolverket.
- Skolverket (2011b): *Bedömningsanvisningar till ämnesprov Svenska/Svenska som andraspråk 2011. Delprov A. Att läsa och förstå*. Stockholm: Skolverket.
- Skolverket (2012a): *Med fokus på läsande: analys av samstämmigheten mellan svenska styrdokument, ämnesprov i svenska och PIRLS 2011*. Stockholm: Skolverket.
- Skolverket (2012b): *Ämnesproven 2011 i grundskolans årskurs 9 och specialskolan årskurs 10*. Stockholm: Skolverket.
- Snow, Catherine E. (red.) (2002): *Reading for Understanding: Toward a R&D Program in Reading Comprehension*. Washington DC: RAND Corporation.
- SOU 2012:65. *Läsandets kultur. Slutbetänkande av Litteraturutredningen*. Stockholm: Fritzes.
- Tengberg, Michael (under utgivning): *Konstruktion och bedömning av förmågan att läsa (och förstå) skönlitterär text: Två nedslag i det nationella provets läsförståelsedel, åk 9. Educare*.
- Utdanningsdirektoratet (2010): *Rammeverk for nasjonale prøver*. http://www.udir.no/Upload/Nasjonale_prover/2010/5/Rammeverk_NP_22122010.pdf?epslanguage=no Hämtad 2013-05-27.
- van Steensel, Roel, Oostdam, Ron & van Gelderen, Amos (2012): *Assessing reading comprehension in adolescent low achievers: Subskills identification and task specificity. Language Testing 30(1), 3–21*.
- Wedin, Åsa (2010): *Bedöma eller döma? Språkbedömning och lästest i grundskolans tidigare år. Pedagogisk Forskning i Sverige 15(2), 102–114*.
- Wiliam, Dylan (2010): *What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. Review of Research in Education 34, 254–284*.
- Östlund-Stjärnegårdh, E. (2010): *Skrivutveckling speglad i prov. I Madeleine Ellvin & Lena Manderstedt, red: Växtkraft: Om svenskämnets möjligheter, s 133–147*. Stockholm: Svensk lärarföreningen.