

Konstruktion och bedömning av förmågan att läsa (och förstå) skönlitterär text. Två nedslag i det nationella provets läsförståelsedel, åk 9

Michael Tengberg

This article discusses the validity of reading comprehension measurements in the 9th grade national assessments of Swedish. It argues for the necessity of establishing suitable construct definitions of reading comprehension before any valid inferences are to be made from students' test results. Theory of assessment and theory of literary response are drawn upon in order to analyze, on item level, the way in which good reading comprehension is represented in the test. The results demonstrate two different types of shortcomings related first of all to inadequate construct definitions. The first problem is the narrow perspective from which acceptable interpretations in student's literary responses are defined. The second problem is the priority given to quantitative over qualitative facets when assessing students' constructed-responses.

Keywords: domain, national assessments, reading skills, reading, validity

Michael Tengberg, universitetslektor, Karlstads universitet
michael.tengberg@kau.se

Inledning

Debatten om ungas läsning och om skolans läsundervisning har på senare år präglats alltmer av vad man skulle kunna kalla ett mätperspektiv. Resultaten från internationella jämförande mätningar av elevers läsförmåga i PISA och PIRLS är högfrekventa referenser både i samhällsdebatten och i skolforskningen (Pettersson, 2008). I den nyliberala utbildningspolitiken har kunskapsmätningar i form av tidigare betyg och fler nationella prov blivit ledstjärnor för framtidens skola. När det gäller själva utformningen av de nationella proven menar vidare Skolinspektionen att utrymmet för tolkning i proven bör minska för att kunna säkra möjligheterna till en likvärdig bedömning (Skolinspektionen, 2011). Viljan att mäta det som är komplext och kontext-

bundet ställs därmed mot kravet att det som mäts ska vara enkelt och just mätbart.

Att beskriva det styrsystem som omgärdar läsningen i skolan som nyliberalt är förstås i sig inte ideologiskt neutralt. Tvärtom är det alltid en politiskt värdeladdad handling att belägga andras dokument eller handlingar med ideologiska etiketter. Utan att fördjupa mig vidare i den här frågan vill jag som första anhalt i den här texten utan omsvep bekänna att analysen är författad utifrån ett perspektiv som bejakar behovet av ett bedömningssystem och av nationella prov i skolan. Samtidigt menar jag att bedömningssystemets begränsningar vad gäller möjligheten till prediktioner om individers och gruppernas läsförmåga är en fråga som behöver analyseras närmare.

Kursändringen mot en allt större tilltro till det s.k. mätperspektivet riskerar också att leda till två improduktiva och därför djupt problematiska hållningar inom ramen för forskning och utbildning som relaterar till svenskundervisningen i skolan. Den ena problematiska hållningen gör gällande att det inte går att mäta kunskaper, att summativa bedömningar av elevers kunskaper per definition är skadliga och att de som ägnar sig åt skolstatistik inte själva kan något om skolans ämnesinnehåll eller om skolan som arbetsplats för elever och lärare. Ett sådant synsätt företräds ofta av dem som själva inte begagnar sig av (och kanske heller inte behärskar) statistiska forskningsmetoder. Den motsatta, och lika problematiska, hållningen innebär en generell övertro på implikationerna från skolstatistiken, och yttrar sig bland annat hos debattörer som förutsätter att lösningen på våra skolproblem finns i Finland (Enkvist, 2011), att mer läsning automatiskt gör bättre läsare (SOU 2012:65), eller att svensklärare är oförmögna att sätta rättvisande betyg på sina elever (Skolinspektionen, 2011, s. 5). Det synsättet företräds av debattörer som söker enkla lösningar eller av forskare som är duktiga på statistik, men som saknar närmare kännedom om skolans ämnesmässiga och sociala verklighet.

När politiker och tjänstemän sätter allt större tilltro till resultat från jämförande kunskapsmätningar blir det desto viktigare för forskare att granska premisserna för själva mätandet (Forsberg & Lundahl, 2006). Dels blir det angeläget att på ett övergripande plan undersöka hur proven förhåller sig till testteoretiska grundförutsättningar. Dels blir det angeläget att från ett ämnesperspektiv problematisera de begreppsdefinitioner som proven explicit eller implicit tillämpar, d.v.s. vilken syn på fenomen som läsning och läsförmåga som proven representerar.

I den här artikeln vill jag pröva hållbarheten i det nationella bedömningsinstrumentet av elevers förmåga att läsa (och förstå) skönlitterär text genom att analysera hur läsförmåga som fenomen konstrueras av uppgifter i det nationella provets läsförståelsedel (delprov A) i årskurs 9. Eftersom jag här intresserar mig för läsning av skönlitterär text sätter jag ordet 'förstå' inom parentes. Från litteraturvetenskapligt och litteraturdidaktiskt perspektiv betraktas inte konkret förståelse som ett överordnat mål för den litterära läsprocessen på samma sätt som man (åtminstone intuitivt) tänker sig att man ägnar sig åt att 'förstå' innehållet i en nyhetsartikel i morgontidningen. Snarare ställer den litterära läsprocessen upplevelse, fantasi, kreativitet och främmandegöring i utsikt, erfarenheter som hellre är att betrakta som performativa än konstativa till sin karaktär (Attridge, 2004, s. 5ff). Vad innebär detta för prov och provuppgifter som avser att testa elevers förmåga att läsa (och förstå) skönlitterär text? I vilken mån förhåller sig provuppgifterna till centrala teorier om relationen mellan läsare, text och tolkning?

Provinstitutionens sätt att konstruera förmåga att läsa (och förstå) skönlitteratur diskuteras här genom två nedslag i för provet teoretiskt och praktiskt problematiska aspekter. Min tes är att dessa testteoretiska och textteoretiska problemområden måste hanteras om man från statligt håll ska kunna framhärda i att det är en legitim verksamhet att mäta elevers förmåga att läsa skönlitterära texter. I vågskålen ligger förstås en väsentlig dimension av elevers betyg i ämnet svenska och därmed en långtifrån försumbar del av elevernas möjligheter till framtida utbildningar, yrken och livsval. Det är alltså fråga om en myndighetsutövning av stor betydelse för individen som därför måste hanteras på ett vederhäftigt sätt. Problemområdena ska förstås som en relation mellan den valda provtexten, frågeformuleringar i provet samt de för provet angivna bedömningsanvisningarna. De två områden som berörs i undersökningen är:

1. En alltför snäv tolkningsram i förhållande till den skönlitterära texten när det gäller att definiera acceptabla svar på provuppgifter
2. Fokus på kvantitativa aspekter av svar på provuppgifter istället för på kvalitativa aspekter

Avgränsningar

Dessa två aspekter av proven diskuteras med utgångspunkt i några exempel från provuppgifter till två kortprosatexter som funnits med i de senare årens ämnesprov. Urvalet av uppgifter gör inte anspråk på att vara representativt för alla uppgifter till skönlitterära texter utan motiveras snarare av att just dessa uppgifter exemplifierar en intressant problematik som jag menar behöver diskuteras. Exempelen är hämtade från ämnesprov som genomfördes före 2013 och som därmed relaterar till kursplanen i svenska i Lpo94. Införandet av en ny kursplan i Lgr11 har naturligtvis inverkat på definitionen av fenomenet läsförmåga och rimligen också på utformningen av ämnesproven. Ämnesprov i årskurs 9 enligt den nya kursplanen genomfördes emellertid först 2013 och fanns inte tillgängliga när föreliggande analys genomfördes. Kursplanerevideringens betydelse för provet uppmärksammas följaktligen inte i den här artikeln. Det är dock fortfarande samma arbetsgrupp som konstruerar ämnesproven¹, varför det ändå ter sig relevant att syna provkonstruktionen närmare.

Validitet och testosäkerhet

Konstruktionen av den förväntade läsförmågan i ett nationellt prov hänger ihop med frågan om tillförlitligheten i själva bedömningsinstrumentet. Mäter provet det som det avser att mäta? En grundläggande testteoretisk utgångspunkt är att själva konstruktionen av ett test måste börja med en definition av den *kunskapsdomän* (Koretz, 2008, s. 19) som provet avser att ta ett mått på. När vi prövar elevers kunskaper är vi ju inte enbart intresserade av hur väl de behärskar just provuppgifterna. Istället vill vi kunna generalisera resultaten till att utgöra mått på ett vidare kunskapsområde (Nuttall, 1987). Wiliam understryker att “unless assessment design begins with construct definition, then establishing the inferences that may validly be drawn from the assessment outcomes becomes difficult, if not impossible.” (Wiliam, 2010, s. 264) I det här fallet bör en sådan definition utgå från vad kursplanen i svenska säger om elevers förväntade läsförmåga i årskurs nio. Elevers prestationer på det nationella provets läsförståelsedel ska alltså erbjuda en rimlig representation av deras kunskaper inom domänen läsning, så som den definieras i kursplanen. Ett problem härvidlag är förstås att kursplaner – detta gäller i synnerhet svenska kursplaner efter 1994 – ofta är kortfattade och mångtydi-

¹ Nationella provgruppen vid Institutionen för nordiska språk, Uppsala Universitet

ga dokument, som dels förutsätter ämnesrelevanta inferenser och dels får olika betydelser givet vilka inferenser som görs (jfr Hultin, 2006). Wiliam menar att begreppsdefinitioner av det här slaget måste avgöras av ämnesexperter på området innan man konstruerar prov för att mäta en viss kunskapsdomän. Annars är det troligt att det som är praktiskt eller billigt att mäta får större inflytande över provdesignen än det som på god grund kan anses viktigt att mäta (Wiliam, 2010, s. 260).

En annan testteoretisk utgångspunkt handlar om provets validitet. Ofta definierar vi slentrianmässigt validitet som en egenskap i mätningen, d.v.s. det att ett prov, eller en undersökning, mäter det som den avser att mäta. Men faktum är att prov i sig själva inte avser att mäta någonting – de mäter bara det de mäter helt enkelt. Däremot drar människor slutsatser av olika slag baserade på proven (exempelvis prognoser om en elevs förutsättningar att klara en viss utbildning) och dessa slutsatser kan med utgångspunkt i ett visst provresultat vara mer och mindre valida. Detta brukar benämnas *begreppsvaliditet* och har att göra med i vilken utsträckning ett visst testresultat kan användas för att dra slutsatser som får utbildningsmässiga eller andra konsekvenser. En gängse definition formulerad av Messick lyder: "Validity 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." (Messick, 1989, s. 13) Det som behöver valideras är alltså, enligt Messick, tolkningen eller betydelsen av ett mätresultat, d.v.s. i vilken mån vi anser att mätresultatet fungerar som en indikator på en viss egenskap hos en individ eller en grupp.

Messick tar också upp två olika slags hot mot validiteten. Å ena sidan kan ett test vara konstruerat så att det försummar att mäta väsentliga delar av den kunskapsdomän som man hoppas kunna generalisera resultatet till. Låt säga att vi vill pröva elevers läsförståelse genom att låta dem svara på flervalsfrågor om några texter som de får läsa. På det sättet blir det möjligt exempelvis att få reda på om de kan lokalisera information i texter och göra enklare analyser/tolkningar av textens innehåll och form. Men det är tveksamt om ett sådant prov kan säga något om elevernas förmåga att relatera det lästa till egna tankar och erfarenheter eller om deras förmåga att värdera texter av olika slag.² Detta kallar Messick *construct under-representation*. Ett test kan också vara konstruerat så att det mäter aspekter som egentligen inte ingår i den valda kunskapsdomänen och som det därför inte borde mäta. Vi kan

² De målsättningar som exemplifieras här är hämtade från kursplanen i svenska i Lpo 94.

tänka oss ett läsförståelseprov där själva frågeformuleringarna innehåller ord som en stor andel elever inte är bekanta med. Risken är då att eleverna inte får tillfälle att visa sin förståelse av den lästa texten, eftersom de missförstår eller helt enkelt hoppar över frågan. Provet har på så vis prövat dem på en annan typ av kunskaper än vad som var avsikten. Den här sortens problematik kallar Messick *construct-irrelevant test variance* (Messick, 1989, s. 34).

Hur frågor formuleras, vilka texter som väljs och vilka svar som anses acceptabla får därmed stor betydelse för vad man egentligen kan säga sig pröva i ett läsförståelseprov. Exempel på svensk forskning om läspröv finns i Wiksten Folkeryd och af Geijerstam (2008) samt af Geijerstam och Wiksten Folkeryd (2013) som undersökt vilken slags läsning som konstrueras i PIRLS-undersökningarna givet de testinstrument och de bedömningsprocedurer som används. Genom att särskilt studera de svar som bedöms som felaktiga har de bl.a. konstaterat att felen oftare handlar om att eleven missförstått frågan än om att den missförstått den lästa texten. Wedin (2010) har i en analys av lästestet ”Vilken bild är rätt?” för grundskolans tidigare år visat på liknande problem. Frågorna i testet kan lätt misstolkas och leda till indicer om bristande läsförmåga trots att problemet alltså egentligen är något annat. Lästest av det här slaget riskerar därmed att ge både lärare, föräldrar och beslutsfattare vilseledande information om hur det egentligen står till med elevernas läsförmåga.

Text, läsare och tolkning

Sett från ett textteoretiskt och receptionsteoretiskt perspektiv befinner sig den skönlitterära texten alltid i ett öppet förhållande till historiska och kulturella sammanhang, vilket innebär att någon given tolkning av ett verk inte existerar (Malmgren, 1983, s. 14). Med tolkning menar jag här allt aktivt förståelsearbete eller nödvändigt referensskapande som läsaren ägnar sig åt. På ett principiellt plan blir det problematiskt att försöka dra en gräns mellan att ”omedelbart förstå” och att tolka (jfr Agrell, 2009, s. 36). Även den till synes omedelbara förståelsen bygger ju på en uppsättning förväntningar eller konventioner som blir verksamma i en lässituation men inte i en annan. Inom läsforskningen används begreppet tolkning emellertid ofta för att beteckna en viss läsart, ett reflekterat och analytiskt betydelseskapande som skiljer sig från den initiala, immersiva läsningen (Appleyard, 1991, s. 121f). Oavsett vilket måste tolkning betraktas som en socialt bestämd företeelse och före-

ställningen om att det finns 'korrekta tolkningar' av texter avvisas samfällt i textteorin som en förlegad myt.

Det råder emellertid inte någon enighet inom receptionsforskningen om hur premisserna för läsarens, textens och tolkningens intrikata förhållande ska definieras. I ett svenskt litteraturdidaktiskt perspektiv antar man allmänt ett *transaktionellt* synsätt på relationen mellan text och läsare, ofta med hänvisning till Louise Rosenblatt (1978). Nu definieras denna transaktion visserligen på olika sätt av olika forskare (se Mehrstam, 2009), men i dess mest renodlade form betraktar man från ett transaktionellt perspektiv inte ens den litterära teckenstrukturen, d.v.s. det läsaren antas respondera på, som någonting som finns nedlagt i texten utan snarare som något läsaren själv realiserar eller erfår under läsningens gång. Texten blir till text i det ögonblick en läsare betraktar den som sådan och konstitueras på ett unikt sätt i varje enskild läsning. Rosenblatt skriver: "He [läsaren] evolves an impalpable experiential structure which he apprehends as the poem or novel or play." (Rosenblatt, 1978, s. 50)

Men även i ett *interaktionellt* perspektiv, där textens struktur antas vara "intersubjektivt tillgänglig [för] analys och diskussion" (Agrell, 2009, s. 35), betraktas tolkningsakten som ofrånkomligt rotad i ett lässammanhang och bunden till den enskilde läsarens minnen, erfarenheter, övertygelser och känslor. Detta komplicerar givetvis alla antaganden om bedömningsrationallitet och om adekvata och likvärdiga bedömningar av elevers "textförståelse". I en provsituation likt den som det nationella provet utgör får den konkreta läsaktens inramning i form av lärares och texthäftets instruktioner, provfrågornas formuleringar och inte minst hela situationens provkaraktär naturligtvis betydelse för elevernas läsning och tolkning. Detta måste tas hänsyn till i såväl bedömning av elevernas prestationer som i analysen av vilket slags läsförmåga provet mäter. I det följande lånar jag detta synsätt för att illustrera problemen med de avgränsningar av tolkningsutrymmet som provinstitutionen gör i syfte att konstruera ett rationellt och likvärdigt bedömningsinstrument.

En alltför snäv tolkningsram

I läsförståelsedelen 2011 ingick en kortprosatekst med titeln "Den elektriska flickan" (Nilsson, 1998). I texten berättas det om en ovanlig flicka, som redan i tidig ålder har speciella egenskaper. Hon är självlysande, folk får stötar när de tar i henne och hon kan baka bullar genom att bara hålla degklumpar i

händerna. Texten är knappt 1200 ord lång och i några kort skisserade scener får läsaren följa flickans liv från födseln fram till hennes död då hon, efter vad man förstår, brinner upp i förening med en ung man som, antydningssvis, också han är elektrisk. I berättelsens inledning får vi veta att det hela utspelar sig på den tiden då elektriciteten var ett nytt fenomen ”någon gång i början av seklet i en liten by i skogstrakterna i Blekinge.” Så här berättas det sedan om flickans första levnadsår:

Just när strömmen drogs till denna by föddes en flicka. Hon var ett ilsket barn som skrek mycket och länge. Hennes far var elektriskt intresserad. Han var den förste i byn som köpte lampa och han hade på ett ungefär förstått hur sladdarna skulle kopplas för att det verkligen skulle bli ljus.

Flickebarnet var ilsket och skrek och skrek. Hennes far beställde då från Claes Ohlsson i Insjön en elektrisk motor. Den byggde han in i vaggan och genom en fiffig koppling skulle vaggan fås att vipa fram och tillbaka av sig självt. Det skulle bli en stor lindring för mor.

Det gick inte riktigt som han hoppats. Ljuset i taket började genast flacka. Allt ljus i hela byn började flämta. Men detta med elektricitet var så nytt att ingen kom sig för att klaga. Och vaggan vibrerade snarare än vaggade och det surrade om den som en spinnande katt. Men flickan somnade gott till darnningarna och det dova surret.

Efter två år hade hon vuxit till sig. Vaggan hade gjort sin tjänst och ställdes undan på vinden. Det var ungefär då som en inspektör från kraftbolaget dök upp. Man hade blivit av med ström i ledningarna på vägen, sa han. Lika mycket ström som hela järnbruket i Kallinge gjorde av med hade spårlöst försvunnit, men de senaste dagarna, troligen efter hans inspektion hos alla abonnenter, verkade problemet oförklarligt ha lösts, tillade han nöjt.

Det var något märkligt med flickan, det såg föräldrarna genast. Hon kunde inte röra en katt utan att pälsen ställde sig rakt upp och den for iväg som om den fått ett slag på nosen. Så katter gick ur vägen för henne.

Det kunde också spraka om henne och ibland om nätterna var hon liksom självlysande. Men detta var ju deras första barn och hon var som sagt ovanligt ilsken av sig.

Konstruktion och bedömning av förmågan att läsa (och förstå) skönlitterär text.
Två nedslag i det nationella provets läsförståelsedel, åk 9

Flickan hade energi på något vis, det var alla ense om. Hon kunde arbeta i slåttern tre dagar och tre nätter i sträck utan att äta eller dricka. Hon gick som en liten maskin över fälten. Men vilken bonde klagar över en sådan flicka?

I provet, som innehåller fem olika frågor till den här texten samt ytterligare ett antal frågor till andra texter, ställs följande fråga:

En man från kraftbolaget skulle ta reda på varför det hade försvunnit så mycket ström i ledningarna. a) Hur hade strömmen försvunnit? b) Varför hittade han inget fel? (Nivå 3)

Markeringen ”Nivå 3” anger att elevens svar kan bedömas upp till och med nivå 3 som är den högsta nivån och innebär att eleven lämnar ”utförligare svar där flera aspekter på uppgiften efterfrågas.” (Skolverket, 2012, s. 45) (Detta diskuteras mer utförligt längre fram i artikeln.) I någon mening handlar detta alltså om att definiera en frågas svårighetsgrad. Svar som inte anses fullständiga kan dock bedömas ligga på nivå 1 eller nivå 2, med hänvisning till anvisningarna i bedömningsmallen. En frågas svårighetsgrad avgörs dels genom provinstitutionens egen bedömning, dels i den statistiska utprovningen. I den medföljande bedömningsmallen ges följande anvisning för lärarnas bedömningar av elevsvar:

Uppgifter	0-svar	Nivå 1	Nivå 2	Nivå 3
5. En man från kraftbolaget skulle ta reda på varför det hade försvunnit så mycket ström i ledningarna. a) Hur hade strömmen försvunnit? b) Varför hittade han inget fel? (Nivå 3)	a) den hade spårlost försvunnit i ledningarna på vägen. a) all ström ljus och elektricitet drog sig till flickan. a) för att flickan var orsaken. b) för att han inte misstänkte att en liten flicka kunde innehålla all ström.	[Eleven har svarat rätt på fråga a). Strömmen användes till att gunga vaggan som barnet sov i eller att det var elmotorn som drog ström.]	[Eleven har svarat rätt på fråga b). Felet har försvunnit eftersom vaggan inte längre används.] för att flickan hade slutat sova i vaggan.	[Eleven har svarat rätt på både fråga a) och b). Kombinationen av laddningen i vaggan och att den just blivit undanställd.] När pappan kopplade motorn till vaggan då drog den en massa energi och det var därför strömmen försvann. När flickan blev två år och vaggan ställdes undan användes inte motorn som var kopplad till vaggan så därför hittade han inget fel.

I Skolverkets årliga rapport om det nationella provets genomförande anges att den ovan citerade frågan om varför strömmen hade försvunnit var den fråga i delprov A som flest lärare hörde av sig om efter provets genomförande 2011 (Skolverket, 2012, s. 46). Lärarna menade att provinstitutionen bedömt felaktigt när den ”underkänner svar som antyder att det var flickan som var orsaken till att elen hade försvunnit” (Ibid.). Även om Skolverket samlade in ett urval av 1590 elevresultat och explicit diskuterar just den här provfrågan på flera sidor i rapporten, redovisar man inte hur eleverna har svarat på den, utan endast att ett stort antal lärare varit kritiska till bedömningen. Det är dock rimligt att tänka sig att denna kritik delvis är baserad på hur

dessa lärares elever svarat. Inom ramen för en annan undersökning (ännu opublicerad) har jag tillsammans med en kollega låtit drygt tvåhundra sjunde- och åttaklassare läsa "Den elektriska flickan" och besvara bl.a. den här frågan i en provliknande situation. I dessa elevsvar är uppfattningen att det är flickan som är orsaken till att elen försvunnit, d.v.s. att flickan har fått strömmen i sig, mycket vanlig (55 %). Eftersom resten av berättelsen bygger vidare på det faktum att hon är elektrisk – kattens päls reser sig när hon rör dem, hon är självlysande, ovanligt energisk etc. – är det kanske heller inte en så konstig tolkning. Man kan fråga sig vad själva episoden med vaggan annars fyller för funktion i texten. Provinstitutionen ger dock svar på tal i rapporten, d.v.s. svar på lärarnas kritik.

Bedömningsmallen underkänner svar som antyder att det var flickan som var orsaken till att elen hade försvunnit, eftersom det som gjorde att strömmen försvann från början är pappans felkoppling av motorn. Detta gjorde att vaggan tog onaturligt mycket ström och den hade slukat elen oavsett om flickan sov i vaggan eller inte. Strömmen kommer tillbaka när vaggan ställs undan på vinden. Om det var flickan som tog elen skulle problemet kvarstå trots att vaggan blev undanställd. (Skolverket, 2012, s. 46)

De elevsvar jag studerat uttrycker dock inte primärt att flickan skulle ha "fått i sig elen" oberoende av vaggan, utan att hon legat i vaggan och på det sättet fått elen i sig, eller blivit uppladdad. Eleverna skriver t.ex.: "Det var flickan som var orsaken, när hon låg i vaggan som pappan kopplat en motor till." Många lärare och elever menar alltså att även om vaggan spelade en roll i elens försvinnande så var poängen ändå att det var flickan som fick den i sig. Provinstitutionen däremot hävdar att det var vaggan i sig som tog onaturligt mycket ström, helt oberoende av flickan. Problemet med provinstitutionens texttolkning är att den inte förklarar hur flickan blev elektrisk, en fördel som elevernas tolkning istället har. Det är frestande att här inflika att en återkommande tendens i läsprovens frågeformuleringar under åren är att instruera eleverna att deras svar ska ha stöd i texten (se exempel nedan). För provinstitutionens egen inferens att vaggan skulle ha dragit lika mycket ström även om flickan inte hade legat i den blir det emellertid svårt att hitta något explicit stöd i texten. Inte heller gör provinstitutionen själv någon ansats till att presentera grunden för sin läsning av texten. En rimligare tolkning synes det därför vara att det är just är genom att ligga i den elektriska vaggan som

flickan 'blir elektrisk'. Märk väl, det står ingenting i texten om att hon *föds* elektriskt laddad.

Min poäng är givetvis inte att försöka avgöra vilken av dessa tolkningar som är mest kvalificerad. Istället vill jag lyfta fram problematiken i provinstitutionens snäva sätt att definiera det korrekta sättet att förhålla sig till en i sammanhanget intressant tolkningsfråga. Detta problem fördjupas ytterligare när provinstitutionen som svar på lärarnas kritik ser sig tvingad att försvara den egna läsningen med inferenser som i sammanhanget ter sig lösryckta, eftersom de inte beläggs med stöd i texten. Till provinstitutionens försvar bör det understrykas att konstruktionen av proven ofta föregås av omfattande utprovningar i stora elevgrupper och med referensgrupper av lärare. Frågors lämplighet, exakta formulering och svårighetsgrad avpassas på det sättet till den valda målgruppen och till den specifika skolkontexten. Samtidigt visar exemplet ovan på ett mer generellt problem, som inte enbart handlar om den specifika frågeformuleringen. Provinstitutionen väljer medvetet att dra en snäv gräns runt de svarsalternativ som anses acceptabla och definierar därmed på ett smalt sätt innebörden i att antingen 'förstå' eller 'inte förstå' centrala aspekter av de valda provtexterna.

Sedan våren 2010 genomför Skolinspektionen s.k. omrättningar av de nationella proven i syfte att avgöra i vilken mån proven möjliggör en likvärdig bedömning och betygsättning. Baserat på de egna omrättningarna, som visar på omfattande och huvudsakligen negativa avvikelser från de ursprungliga rättningarna³, föreslår Skolinspektionen att tolkningsutrymmet minskas i bedömningsanvisningarna. I första hand gäller kritiken de s.k. skrivproven, i vilka elever ska skriva egna längre texter på ett valt ämne, men för årkurs 9 är de negativa avvikelserna nästan lika stora på läsprovet som på skrivprovet (Skolinspektionen, 2011, s. 75, 2012, s. 13). En anpassning till de principer Skolinspektionen förordar skulle alltså kunna leda till att provinstitutionen gör allt större ansträngningar för att minska tolkningsutrymmet – och därmed utrymmet för lärares godtycke – i bedömningen av enskilda uppgifter. Men som den här analysen visar finns det ett motsatt problem, som inte handlar om att utrymmet för lärarnas godtycke är för stort, utan att provinstitutionens godtycke ifråga om texttolkningen sätter alltför snäv gräns för konstruktionen av fenomenet förmåga att läsa (och förstå) skönlitterär text.

³ Negativa avvikelser innebär att Skolinspektionen bedömer att elevens prestation motsvarar ett lägre betyg än vad ursprungs rättaren bedömt.

Konstruktion och bedömning av förmågan att läsa (och förstå) skönlitterär text.
Två nedslag i det nationella provets läsförståelsedel, åk 9

Kvantitet eller kvalitet?

Det andra exemplet på den problematiska konstruktionen av god läsförmåga har att göra med nivågraderingen av provfrågor och svarsalternativ i bedömningsanvisningarna. Här anges tre möjliga 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 (Skolverket, 2012, s. 45).

Som redan nämnts kan en fråga som medger svar på nivå 3 även möjliggöra svar på nivå 1 eller nivå 2. I bedömningsanvisningarna definierar provinstitutionen nivåerna i förhållande till varje enskild provfråga. Uppdelningen ska ses som kvalitativ eftersom andelen frågor som eleven lyckas besvara på nivå 2 respektive nivå 3 blir avgörande för chanserna till de högre betygen VG och MVG. Man bör alltså dra slutsatsen att det inom ramen för provsammanhanget betraktas som ett uttryck för mer avancerad läsförmåga att svara på nivå 3 än på nivå 2 eller nivå 1.⁴ Synar man bedömningsanvisningarna närmare blir det emellertid svårt att hålla med om detta, både utifrån ett testteoretiskt och från ett textteoretiskt perspektiv.

I läsförståelsedelen år 2006 ingick i texthäftet en kortprosatext med titeln ”Liten julnovell” (Göransson, 1998). Den handlar om Bettan och Benny, som förmodas vara väldigt fattiga, eventuellt uteliggare. Inför julafton har Bettan samlat burkar nog för att ge dem råd att äta på IKEA. Efter middagen gömmer de sig inne på lagret tills varuhuset har stängt och bygger sedan upp ett litet vardagsrum med soffa, bokhyllor och säng och spenderar julnatten därinne. Det är Bettan som är den drivande och som instruerar Benny vad de ska göra. En av provfrågorna till texten lyder:

⁴ I bedömningsanvisningarna anger provinstitutionen exempelvis att nivå 3 motsvarar ”bästa tänkbara svar” (Skolverket, 2006, s. 3).

Vilken beskrivning passar bäst in på Bettan? Motivera ditt svar. Ditt svar ska ha stöd i texten. (Nivå 3)

- | | |
|--|--|
| ☐ ensam och kriminell | ☐ fattig och påhittig |
| ☐ rädd och ensam | ☐ planlös och fattig |

För ett korrekt svar på nivå 1 ska eleven, enligt bedömningsanvisningarna kryssa i alternativet fattig och påhittig. För nivå 2 ska eleven dessutom motivera ena ledet väl, t.ex. att man vet att hon är fattig för hon samlade ju burkar för att ha råd med maten. För nivå 3 ska eleven ha motiverat båda leden väl, t.ex. genom att lägga till att Bettan är påhittig för hon kommer ju på hela idén med att de kan bygga ett vardagsrum inne på IKEA. Det spelar dock ingen roll vilket av leden eleven motiverar för att få nivå 2, poängen är att motivera ett av dem. Lägg märke till att det inte finns någon information till eleven om hur svaret kommer att bedömas, t.ex. att båda leden måste finnas med för att få nivå 3. En elev som anger flera goda skäl med stöd i texten för varför Bettan är påhittig kan heller inte få lika hög bedömning som en elev som korthugget anger ett skäl för vart och ett av leden.

I samma prov finns ytterligare en fråga om Bettan.

Bettans julklapp till Benny (ett mål mat och ett hem för en natt) är planerad långt i förväg. Hur förstår man det av texten? (Nivå 2)

Här ska eleven formulera sitt svar helt självständigt. Av bedömningsmallen framgår att om eleven förklarat planeringen till en av julklapparna (alltså antingen maten eller hem för en natt) så ger det nivå 1, och om planeringen för båda julklapparna motiveras så ger det nivå 2. Återigen, ordningen spelar ingen roll utan det är hur många delar av frågan som berörs i svaret som blir avgörande för bedömningen.

Det provinstitutionen fokuserar på i de här två frågorna är uppenbarligen inte kvalitativa utan kvantitativa skillnader mellan elevsvaren. Bedömningen skiljer inte en mer elementär från en mer avancerad läsförmåga, vilket såväl

styrdokumenten och betygssystemet som det nationella provsystemets definitioner av olika svarsnivåer förutsätter. Istället för att skilja på kvalitativt olika nivåer av svar bedöms hur många led i frågan som eleven motiverat. Även om provfrågorna kan tyckas medge kvalitativa bedömningar så premierar bedömningsanvisningarna snarare mer fullständiga svar än svar av högre kvalitet. Detta är ett drag som återkommer i många av läsförståelsedelens provfrågor. Dessutom anges inte i frågan vilka kriterier som svaret kommer att bedömas utifrån.

Utöver detta kan man diskutera rimligheten i den angivna nivåindelningen. Nivå 1 anges vara kortsvar som kan utläsas direkt ur texten. Men att Bettan i "Liten julnovell" är just fattig och påhittig står inte uttryckligen i texten. Tvärtom förutsätter en sådan bedömning att eleven väger samman Bettans olika handlingar i berättelsen. Av de drygt tvåhundra sjundeklassare som vi låtit läsa samma text och besvara bl.a. den här provfrågan är det dessutom ytterst få som inte har rätt på nivå 1. Notera att det är ett prov för årskurs 9. Givet att man nu betraktar Bettan som fattig och påhittig så är själva motiveringen till detta – snarare än karaktäriseringen i sig – något som kan utläsas direkt i texten. Att göra denna motivering har betydligt fler elever svårt med, men det är tveksamt om det betyder att de har sämre förmåga att läsa en skönlitterär text. Snarare tycks det peka på att de är sämre tränade i att leta upp stöd för sina läsupplevelser i de lästa texterna och formulera detta skriftligt i en provsituation. Det innebär i så fall att den bedömning (och eventuellt den statistiska utövningen) som legat till grund för nivåindelningen av svaren visserligen skiljt på uppgifternas relativa svårighetsgrad, men att svårighetsgraden inte direkt kan relateras till kvalitativt olika nivåer av läsförståelse utan att den samtidigt mäter något annat. Det innebär förstås att även provet mäter något annat än vad det antas mäta.

Diskussion

Avsikten med den här studien är inte att försöka ge någon heltäckande bild av hur läsförståelseprovet definierar läsförmåga utan att belysa en befintlig problematik inom ramen för provsystemet. Även om de problem som lyfts fram i analysen återkommer i flera provfrågor och i prov från olika år så måste det bli en annan studies uppgift att svara på hur pass utbrett problemet är eller vilka konsekvenser det får för provresultatens sammanlagda tillförlitlighet. Likväl måste det påpekas att om kritiken mot det snäva tolkningsutrymmet i de tillåtna svaren och mot tendensen att premiera kvantitet som ett

uttryck för kvalitet kan anses giltig så finns det risk att provfrågorna, och därmed provet i sig, utgör en dålig representation av den kunskapsdomän som man avser att pröva, nämligen elevers läsförståelse. Undersökningen antyder, med Messicks termer, att provkonstruktionen lider både av construct under-representation och av construct-irrelevant test variance. Både direkt och indirekt utger sig provet för att göra kvalitativ åtskillnad mellan en enklare och en mer avancerad läsförståelse, men det är alltså tveksamt om provuppgifterna tillsammans med bedömningsanvisningarna möjliggör en sådan differentiering. Vidare antyder analysen att även andra aspekter än kvalitet i läsförståelse kan vara begränsande för bedömningen av elevers prestationer.

Testteoretiskt och textteoretiskt menar jag att provet har att förhålla sig till två övergripande problemområden. Det ena är begreppsdefinitionen i sig, som utan tvivel måste föregå varje försök till prövning av elevers läsförståelse. I bedömningsanvisningarna diskuteras visserligen både nationella kursplanemål och för provet fristående betygskriterier (se Skolverket, 2011, s. 1–4). Men hur de senare kriterierna egentligen förhåller sig dels till den nationella kursplanens mål och betygskriterier och dels till den ovan nämnda nivågraderingen i provfrågorna framgår inte. Att definiera kunskapsdomänen läsning inom ramen för det nationella provsystemet borde dels vara en fråga för öppen debatt inom det litteraturdidaktiska forskningsfältet och dels ett ärende för en nationell grupp av experter på området. Hittills har ingetdera varit fallet.

Det andra övergripande problemområdet handlar om att höjda förväntningar på "likvärdighet i bedömningen" riskerar att leda till validitetsproblematik. Om provsystemet väljer att prioritera uppgifter där det finns facit eller tydliga rätt och fel-alternativ framför uppgifter som kräver en professionell bedömares tolkning så kommer det få s.k. wash back-effekter på praktiken. I en kritisk diskussion av kvantitativa indikatorers effekter på de system de mäter har Donald Campbell formulerat följande lag: "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor (Campbell, 1976, s. 49; se även Stobart & Eggen, 2012, s. 3). Med andra ord, när proven tillmätts allt större betydelse för beslutsfattande om utbildningssystemet kommer lärare att anpassa sin undervisning till provuppgifternas karaktär och innehåll. Om läsproven töms på uppgifter av mer komplex karaktär därför att

Konstruktion och bedömning av förmågan att läsa (och förstå) skönlitterär text.
Två nedslag i det nationella provets läsförståelsedel, åk 9

bedömningen av elevernas svar skapar för stort tolkningsutrymme så riskerar läsundervisningen gradvis också att påverkas i den riktningen.

I den mån staten över huvud taget ska ägna sig åt att mäta elevers förmåga att läsa (och förstå) skönlitterär text – jag vill mena att den frågan också bör diskuteras – behöver vi en ämnesteoretiskt och ämnesdidaktiskt grundad diskussion om hur en operationell konstruktion av fenomenen läsning och läsförmåga kan se ut. Förutom att en sådan domänbeskrivning är nödvändig för provens validitetsanspråk kan den också stödja svensklärares reflexion över litteraturredidaktiska vad- och varför-frågor. Dessutom lär den förse litteraturredidaktiska forskare med angelägna, konkreta och praxisnära forskningsuppgifter för en god tid framåt.

Referenser

- Agrell, Beata (2009). Mellan raderna: Till frågan om textens appellstruktur. I Staffan Thorson och Christer Ekholm (Red.), *Främlingskap och främmandegöring: Förhållningssätt till skönlitteratur i universitetsundervisningen* (ss. 19–148). Göteborg: Daidalos.
- Appleyard, Joseph A. (1991). *Becoming a Reader: The Experience of Fiction from Childhood to Adulthood*. Cambridge: Cambridge U.P.
- Attridge, Derek (2004). *The Singularity of Literature*. Oxon: Routledge.
- Campbell, Douglas (1976). Assessing the impact of planned social change. Paper 8. Occasional Paper Series. Dartmouth College. Hämtad från [https://www.globalhivmeinfo.org/CapacityBuilding/OccasionalPapers/08 Assessing the Impact of Planned Social Change.pdf](https://www.globalhivmeinfo.org/CapacityBuilding/OccasionalPapers/08%20Assessing%20the%20Impact%20of%20Planned%20Social%20Change.pdf) 2013-01-05.
- Enkvist, Inger (2011). Därför är finländska skolan så framgångsrik. I *Dagens Nyheter*, 2011-04-06.
- Forsberg, Eva & Lundahl, Christian (2010). Kunskapsbedömningar som styrmedia. *Utbildning & Demokrati*, 15(3), 7–29.
- af Geijerstam, Åsa & Wiksten Folkeryd, Jenny (2013). Fel, men rätt på något sätt. En analys av elevers nollsvar enligt PIRLS-provet. I Gustaf Skar & Michael Tengberg (Red.), *Läsning* (ss. 37–55). Stockholm: Svenskläraryrket/Natur & Kultur.
- Göransson, Mattias (1998). Liten julnovell. I *Dagens Nyheter*, 1998-12-18.
- Hultin, Eva (2006). *Samtalsgenrer i gymnasieskolans litteraturundervisning. En ämnesdidaktisk studie*. Örebro Studies in Education, 16. Örebro universitet, Pedagogiska institutionen.
- Koretz, Daniel (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge MA: Harvard U.P.
- Malmgren, Lars-Göran (1983). *Litteraturreception och litteraturpedagogik: Utgångspunkter för ett forskningsprojekt*. Lund: LU.
- Mehrstam, Christian (2009). *Textteori för läsforskare*. Göteborg: Institutionen för litteratur, idéhistoria och religion.
- Messick, Samuel (1989). Validity. I R.L. Linn (Red.). *Educational Measurement* (3:e uppl., ss. 13–103). New York: Macmillan.
- Nilsson, Ulf (1998). *Fyra berättelser*. Stockholm: Bonnier Carlsen.
- Nuttall, Desmond L. (1987). The validity of assessments. *European Journal of Psychology of Education*, 2(2), 109–118.
- Pettersson, Daniel (2008). *Internationell kunskapsbedömning som inslag i nationell styrning av skolan*. Uppsala universitet, Pedagogiska institutionen.

Konstruktion och bedömning av förmågan att läsa (och förstå) skönlitterär text.
Två nedslag i det nationella provets läsförståelsedel, åk 9

- Rosenblatt, Louise (1978). *The Reader, the Text, the Poem: The Transactional Theory of the Literary Work*. Carbondale & Edwardsville: Southern Illinois U.P.
- Skolinspektionen (2011). *Lika eller olika: Omrättning av nationella prov i grundskolan och gymnasieskolan*. Stockholm: Skolinspektionen.
- Skolverket (2006). *Bedömningsanvisningar: Under ytan*. Stockholm: Skolverket.
- Skolverket (2011). *Bedömningsanvisningar: Eld och lågor*. Stockholm: Skolverket.
- Skolverket (2012). *Ämnesproven 2011 i grundskolans årskurs 9 och specialskolans årskurs 10*. Stockholm: Skolverket.
- SOU 2012:65. *Läsandets kultur. Slutbetänkande av Litteraturutredningen*. Kulturdepartementet. Stockholm: Fritzes.
- Stobart, Gordon & Eggen, Theo (2012). High-stakes-testing – value, fairness and consequences. *Assessment in Education: Principles, Policy & Practice*, 19(1), 1–6.
- Wedin, Åsa (2010). Bedöma eller döma? Språkbedömning och lästest i grundskolans tidigare år. *Pedagogisk Forskning i Sverige*, 15(2), 102–114.
- Wiksten Folkeryd, Jenny & af Geijerstam, Åsa (2008). Textanalyser av ett lästest: Konstruktion av läsförståelse i PIRLS 2006. Konferenspaper presenterat vid *Svenskans beskrivning 30*, Stockholm 10–11 oktober 2008.
- 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.