



KAPET

Karlstads universitets Pedagogiska Tidskrift



Årgång 15, nummer 2, 2019

KAPET

Karlstads universitets Pedagogiska Tidskrift årgång 15, nummer 2, 2019

Karlstads universitets Pedagogiska Tidskrift handlar om pågående forskning inom Pedagogik, Pedagogiskt arbete, Didaktik och Specialpedagogik, huvudsakligen vid Karlstads universitet. Syftet med skriften är att ge en bild av forskningsverksamheten i Karlstad och att bidra till debatt inom utbildningsvetenskaplig forskning och praktik. Bidrag i form av artiklar, recensioner av avhandlingar eller annan aktuell litteratur och information från konferenser är välkomna.

Tidskriften har till och med 2018 som regel utkommit en gång per år. Från och med nummer, 2019:1, har tidskriften till löpande utgivning av artiklar under året.

Bidrag till KAPET kan skrivas på svenska, andra nordiska språk eller engelska och skickas till redaktionssekreteraren. Tidskriften ges ut av Institutionen för Pedagogiska studier vid Karlstads universitet.

Ansvarig utgivare: Patrik Larsson

Redaktör: Valerie Margrain (e-post: Valeire.Margrain@kau.se)

Redaktionssekreterare: Kent Fredholm (e-post: Kent.Fredholm@kau.se)

Adress: Karlstads Pedagogiska Tidskrift, Institutionen för pedagogiska studier, Karlstads universitet, Universitetsgatan 2, 651 88 Karlstad.

e-ISSN: 2002-3979.

Texterna i tidskriften får kopieras för undervisningsändamål under förutsättning att författaren eller författarnas och tidskriftens namn finns med på kopiorna.

© författarna.

Hängning, tveksamma tester och andra goda motiv för vetenskapliga studier om läsning

Michael Tengberg
professor
Michael.Tengberg@kau.se

Abstract

Hur viktigt är det egentligen att lyckas på ett läsprov? Ja, det beror kanske på i vilket sammanhang provet tas. En lika relevant fråga är dock: Hur viktigt är det att provet lyckas fånga elevernas läsförmåga? Är det ena ens möjligt utan det andra? Den här sortens frågor bildade utgångspunkt för ett föredrag som jag gav i samband med min professorsinstallation den 11 oktober 2019. Avsikten var att försöka representera en del av innehållet i den forskning som jag ägnat mig åt. Den här artikeln utgör en något utvidgad version av föredraget och kretsar i första hand kring frågan om tillförlitlighet i läsprov av olika slag. Men det handlar också om hängning som straff för dåligt genomförda läsprov, om kunskapens möjliga dimensioner och om vad vetenskap egentligen är.

Hängning, tveksamma tester och andra goda motiv för vetenskapliga studier om läsning

Sverige har en lång historia av allmän läskunnighet, längre än de flesta andra länders, förmodligen längre än alla andra länders, tack vare kyrkolagen från 1686, som stadgade att alla svenskar skulle ha kristendomskunskap, vilket i praktiken innebar att man var tvungen att kunna läsa. Att kunna läsa blev därför viktigt. Om man inte kunde läsa fick man inte gifta sig. Inte heller fick man ta nattvarden. Och det ville man. Individen hade således sin morot. Men den oerhört framgångsrika alfabetiseringen av Sverige drevs i första hand av två djupt liggande allmänintressen. Det ena var den protestantiska reformationen, som förespråkade individens personliga kontakt med Gud. Genom Gutenbergs uppfinning hade spridningen av Ordet blivit möjlig. Reformationen gjorde den nödvändig. Thomas Götselius skriver att "[r]eformationen på en gång förutsatte och påbjöd alfabetisering" (Götselius, 2013, s. 42). Det andra överordnade intresset som drev på för läs- och skrivkunnighet (även om de bägge inte alltid förutsatte varandra) var den snabbt växande statsförvaltningen i form av kollegier och kanslier, som förutsatte ämbetsmän som behärskade skriften. Klosterskolor ersattes av byskolor och skolor i städerna. Latinet byttes mot folkspråk. Handelsmän tog vara på skrivkonstens fördelar. Bibel, skola, så småningom tidningar och romaner. Kraven på god läsförmåga höjdes efter hand.

Det finns gott om bevarade exempel på hur utländska gäster som kom till Sverige förundrades över den utbredda läskunnigheten. Under 1700-talet bedömer man att över 90 procent av befolkningen kunde läsa. Nu fanns förstås redan på den tiden det fenomen som vi idag i utbildningssammanhang kallar för teaching-to-the-test, vilket innebar att man inför husförhören ofta lärde sig katekestexten utantill snarare än att man faktiskt lärde sig läsa per definition. Den effekten har ju viktiga prov, det gällde då och det gäller idag.

Läsa för att överleva

Ett av de mest intressanta exemplen på livsviktiga läsprov – och faktiskt också ett av de allra tidigaste exemplen på high-stakes-språktest – kommer från 1100-talets

Michael Tengberg: *Hängning, tveksamma tester och andra goda motiv
för vetenskapliga studier om läsning*

England. Det riktade sig till prästvigda och syftade till att bekräfta just deras klerikala identitet. Kung Henrik II, som i flera avseenden var en framgångsrik regent, hade ett bistert förhållande till den katolska kyrkan, vars juridiska makt han kringkurit genom instiftandet av flera lagar. Efter mordet på ärkebiskopen av Canterbury, Thomas Becket, som det misstänktes att kungen varit delaktig i, tvingades Henrik till eftergifter och lät bland annat präster få särskilda juridiska privilegier (*privilegium clericale*).

Om de anklagades för brott kunde de hävda att de var präster och därmed dömas i en särskild kyrklig rätt, vilket renderade betydligt mildare domar. I början räckte det med att man hade tonsur (den särskilda präst- och munkrakningen på huvudet) och prästkappa. Men från 1350-talet och framåt infördes ett särskilt språktest.

På den här tiden var läskunnighet, precis som det var under husförhörstiden i Sverige, detsamma som att kunna läsa högt. Nu hade visserligen prästskolorna runtom i Europa en långt utvecklad tolkningskonst, men för att testa läsförmågan i det här sammanhanget nöjde man sig med högläsning. Vad prästerna skulle göra för att bevisa att de var präster var helt enkelt att läsa en bibelvers, eller rättare sagt en psalmvers, högt. Det här provet var faktiskt i bruk ända in på 1700-talet. Reliabiliteten i testet var möjligen inte alldeles oförvitlig. Det var nämligen en och samma psalmvers som återanvändes i stort sett hela tiden.

Miserere mei, Deus: secundum magnam misericordiam tuam

Et secundum multitudinem miserationum tuarum: dele iniquitatem meam

(Have mercy upon me, O God,

according to thy loving-kindness:

according unto the multitude of thy tender mercies

blot out my transgressions.)

(Psalm 51: 1 ur King James Version)

Ett lyckosamt resultat på detta läsprov, det vill säga att den misstänkte brottslingen korrekt förmådde högläsa texten på latin, kunde alltså göra skillnad mellan att antingen tvingas till ”andlig botgöring” eller att fängslas eller rentav hängas för brottet, beroende på brottets allvar naturligtvis. Psalmversen ifråga kom därför att gå under smeknamnet *The Neck Verse*. Att präster – och även andra skriftlärda – så småningom lärde sig versen utantill för att inte misslyckas på testet blev naturligtvis känt efterhand.

Exempelvis stod pjäsförfattaren Ben Jonson 1598 anklagad för mord i samband med en duell, men frikändes genom att recitera just denna vers. Jonson hade lyckosamt nog under häktningstiden delat cell med en jesuitpräst. Ibland byttes den berömda versen emellertid ut mot en annan psalmvers för att lura brottslingar som ansågs särskilt förtjänta av en dödsdom.

Våghalsiga slutsatser baserat på moderna lästester

Här lockar det förstås att skriva att jag redan som barn fick höra den här historien och att jag därför alltid känt en passionerad längtan efter att få undersöka närmare de två fenomen som går under namnen reliabilitet och validitet i läsförståelseprov. Men det vore en sanning med viss modifiering. Historien har jag fått höra senare.

En del av min tid som forskare har jag i alla fall spenderat med att undersöka premisserna för att kunna testa läsförmåga och bland annat i vilken mån de tester som används, framför allt i skolan, såsom nationella prov och PISA:s läsförståelseprov, ger underlag för att kunna dra slutsatser om elevers läskunnighet. Ibland får dessa provresultat nämligen ganska långtgående konsekvenser, även om provtagaren numer inte nödvändigtvis riskerar livet. Prov med examinerande eller betygsgrundande funktion får exempelvis högst påtagliga konsekvenser för den enskilde eleven. I fallet med PISA-proven kan konsekvenserna snarare gälla ett helt lands utbildningssystem. Detta sker ofta utan att de flesta, vare sig politiker eller allmänhet, eller läsforskare för den delen, verkar känna till vad proven egentligen prövar, eller hur pass tillförlitliga de är.

Under nästan ett decennium ledde sjunkande resultat på PISA-provet till en omfattande krismentalitet ifråga om kvaliteten på den svenska skolan. En rad reformer på utbildningsområdet genomfördes med provresultaten på PISA som huvudsaklig referenspunkt. Skolor mobiliserade resurser för att sätta in åtgärder på samtliga nivåer och somliga föreläsare gjorde sig en förmögenhet på att resa runt och tala antingen om krisen i sig eller om lösningen på den. I de nyligen presenterade resultaten från PISA 2018 visar resultaten för andra gången i rad på en positiv trend. Svenska elevers läsförståelse uppges vara tillbaka på samma nivå som den var 2006 och dessutom signifikant över OECD-genomsnittet (Skolverket, 2019). Återigen tolkar politiker, journalister, debattörer och skolledare tillståndet i världen (eller i alla fall tillståndet i skolan) utifrån ett enda provresultat. Den här gången är reaktionen betydligt

Michael Tengberg: *Hängning, tveksamma tester och andra goda motiv för vetenskapliga studier om läsning*

stillsammare. Den kan närmast beskrivas som en djup suck av lättnad. Skolverkets tjänstemän underlåter förvisso inte att också rapportera de mindre positiva resultaten, som exempelvis att spridningen i resultat mellan elever har ökat. Den allmänna opinionens sammanfattande reaktion på resultaten tycks dock i huvudsak kännetecknas av en lågmäld belåtenhet.

Om resultaten på PISA-proven är viktiga för nationen så har de desto mindre betydelse (egentligen ingen betydelse alls) för den enskilda eleven. Annorlunda är det med de nationella proven i det svenska skolsystemet. Dessa fungerar så att eleven gör flera delprov i varje ämne och får ett sammanfattande provbetyg. Detta provbetyg ges på den sexgradiga betygsskalan från A till F och ska vara vägledande för lärarens betygsättning i ämnet. Hösten 2017 röstade riksdagen för att låta resultaten från de nationella proven väga tyngre vid lärarens betygsättning än vad de gjort tidigare (Bet. 2017/18:UbU5). En relevant fråga i sammanhanget blir därför om man med de nationella provens hjälp verkligen kan sortera elever i de sex kunskapsnivåer som betygsskalan stipulerar.

För den som minns Betygsberedningen 1992 (SOU 1992:86) är frågan om möjliga kunskapsnivåer bekant. I beredningens förslag ingick ett betygssystem bestående av sex nivåer. Från remissinstanserna kom omfattande kritik. På DN Debatt den 18 augusti samma år invände Ference Marton kraftigt mot förslaget:

Att det skulle finnas sex klart skilda nivåer i kunskapsutvecklingen som regelbundet följer på varandra i alla ämnen är en svindlande och – för en betygsberednings vidkommande – behändigt tanke. Det finns dock ett problem. Tanken är fel. (...) [I]vå decenniers forskning har visat att en sådan strängt lagbunden och allmängiltig följd av nivåer av ökande kompetens inte finns och inte heller kan finnas i verkligheten. (Marton, 1992)

Den gången blev det till slut bara tre steg på betygsskalan. Men förslaget återkom ett par decennier senare och idag har vi som bekant sex steg på betygsskalan, som vart och ett antas representera en urskiljbar och från de andra nivåerna skild kunskapsnivå inom vart och ett av skolans ämnen. Ett nationellt prov med stor inverkan på elevens betyg bör alltså, om det alls ska användas, kunna ge tillförlitlig information om hur

Michael Tengberg: *Hängning, tveksamma tester och andra goda motiv för vetenskapliga studier om läsning*

eleven positionerar sig på den sexgradiga skalan. Detta är naturligtvis inte något man kan förutsätta utan måste prövas empiriskt.

Med ett representativt urval av elever som gjort provet kan man använda något som kallas Rasch-modellering (Rasch, 1980). Det är en statistisk metod som innebär att information om hur svåra eller lätta de enskilda uppgifterna är kan kombineras med information om hur duktiga eller mindre duktiga de enskilda eleverna är. Rasch-analysen ger bland annat något som kallas separationsstatistik. Den inkluderar en skattning av i hur många distinkta förmågenivåer ett prov på ett tillförlitligt sätt lyckas särskilja de deltagande eleverna. Med den här metoden visade jag tillsammans med en kollega att det nationella läsprovet i årskurs nio klarade av att med 76 procents säkerhet skilja på 1,78 olika nivåer av läsförmåga (Tengberg & Skar, 2017). Det vill säga provet klarade nästan att tillförlitligt separera de femhundra eleverna som ingick i analysen i två grupper. Kort sagt, det klarade nästan att skilja de lite bättre läsarna från de lite sämre. Att provet skulle klara av att tillförlitligt dela in elever i sex olika kunskapsnivåer är således en ren fantasi.

Därför är det beklämmande att omkring ett hundratusen elever varje år gör det här provet och att deras avgångsbetyg i svenska starkt påverkas av resultatet på provet. Enligt riksdagens beslut 2017 ska provresultatet alltså väga tyngre än tidigare för avgångsbetyget. Staten gör således ett antagande om att läsförmåga kan mätas med ett visst instrument, medan empiriska bevis pekar på att detta antagande inte håller för en närmare prövning. Man frestas referera till Daniel Koretz, en av USA:s främsta experter på storskaliga tester och dess validitet. I sin bok *Measuring Up* (2008) skriver han bland annat om behovet av försiktighet vid tolkning av enskilda testresultat, också då testerna är stora och välutvecklade:

Of the many complexities entailed by educational testing, the most fundamental, and the one that is ultimately the root of so many misunderstandings of test scores, is that test scores usually do not provide a direct and complete measure of educational achievement. Rather, they are incomplete measures, proxies for the more comprehensive measures that we would ideally use but that are generally unavailable to us. There are two reasons for the incompleteness of achievement tests. One, stressed by careful developers of standardized tests for more than half a

Michael Tengberg: *Hängning, tveksamma tester och andra goda motiv för vetenskapliga studier om läsning*

century, is that these tests can measure only a subset of the goals of education. The second is that even in assessing the goals that can be measured well, tests are generally very small samples of behavior that we use to make estimates of students' mastery of very large domains of knowledge and skill. As explained in the following chapter, an achievement test is in many ways like a political poll, in which the opinions of a small number of voters are used to predict the later votes of many, many more people. (Koretz, 2008, s. 9)

Ett testresultat är med andra ord inte detsamma som vad en elev lärt under ett läsår eller en termin, inte ens när resultatet avser ett nationellt prov. Tvärtom utgör det *en indikation på* elevens kunskapsnivå. Det kan förvisso vara en bättre eller sämre indikation, men likafullt är alla enskilda provresultat ofullständiga. Besinning och försiktighet när det gäller slutsatser lyder därför de stående honnörsorden. Lärares främsta tillgång består inte i resultat på nationella prov – hur välutvecklade dessa prov i en eventuell framtid än skulle kunna vara – utan i det faktum att lärare har möjlighet att observera elevers prestationer vid många olika tillfällen och på olika slags uppgifter. Oberoende bedömning är ett viktigt reliabilitetskriterium, men inte det enda. Upprepad bedömning, vid nya tillfällen, med nya uppgifter, är också viktiga kriterier.

Men är det vetenskap?

På en del av min forskande arbetstid har jag alltså försökt granska vad en bestämd avdelning i statsapparaten har för sig och hur pass välgrundade dess argument är för en viss typ av myndighetsutövning. På det viset hoppas jag att jag och mina kollegor har bidragit med *meningsfull* och kanske rentav *användbar* kunskap. En fråga som ofta återkommer till mig är emellertid om detta också kan kallas vetenskap. Är det en vetenskaplig insats att påvisa och påtala brister i ett läsförståelseprov? Låt vara att provet ifråga har stort inflytande över många människors möjligheter till framtida utbildning och yrkesval. Men har det i sig någon egentlig betydelse för den vetenskapliga kvaliteten? Vetenskap om vadå i så fall? Vad är det för grundläggande form av kunskap om naturen eller människan eller kulturen som jag bidrar till här? I bästa fall kanske jag kan drista mig till att försöka hävda att jag genom att studera de mätinstrument som används för att pröva läsförmåga indirekt kan bidra med kunskap om själva läsandet och dess förutsättningar, exempelvis kunskap om de processer som

Michael Tengberg: *Hängning, tveksamma tester och andra goda motiv
för vetenskapliga studier om läsning*

pågår i huvudet på oss när vi läser. Men kanske är det bara testerna själva, de systemgenererade produkterna, som jag egentligen får veta något om?

Jag har funderat en del över detta nyligen. En av sommarens stora läsoplevelser för min del var Richard Feynmans självbiografi *The Pleasure of Finding Things Out* (2005). Feynman fick Nobelpriset i fysik 1965 för sitt arbete inom kvantelektroniken, men han var också en man som gärna kastade sig in i diskussioner om vad vetenskap är och vad det inte är, inte sällan med generaliserande anspråk, vilket gör flera av hans böcker relevanta för en mycket bredare allmänhet än enbart gruppen universitetsprofessorer i fysik. Feynman företrädde en sträng definition av vetenskap, men även om han använder det engelska ordet *science* syftar han inte enbart på naturvetenskap. *Science* är för Feynman forsknings- eller *undersökningsområden* där man kan etablera vad han kallar *säker kunskap*, det vill säga objektiva och beräkningsbara fakta. Det vill säga områden där man kan undersöka, fastställa och sedan *veta något* med sådan säkerhet att man också kan förutspå exakt hur ett fenomen kommer att bete sig i en framtida händelse och dessutom få dessa förutsägelser bekräftade i nya studier. Det, menar Feynman, är vetenskap. Att verkligen *veta* något, skriver han på ett ställe, att undanröja alla tveksamheter och möjliga alternativa tolkningar av ett experiment, kräver väldigt hårt arbete.

En annorlunda definition av det vetenskapliga arbetet företräder Karl Popper, som i sin bok *Conjectures and Refutations* (1963) skriver om induktionens logik, det vill säga det faktum att vi tenderar att dra slutsatser om lagbundenhet baserat på ofta upprepade observationer. Vetenskapliga observationer, menar Popper, syftar dock inte till att generera teori, utan till att förkasta eller kritisera befintlig teori. Det är genom att antingen lyckas falsifiera en teori eller misslyckas med att falsifiera den och därmed stärka den som vi producerar säkrare kunskap. Inte ens tusen observationer av vita svanar kan bevisa att det endast finns vita svanar, medan observationen av en enda svart svan bevisar att alla svanar *inte* är vita.

Om staten, som i mitt fall ovan, eventuellt baserat på tidigare forskning eller på en serie antaganden kopplade till praktiska erfarenheter, förutsätter att läsförmåga kan mätas med ett visst instrument, och jag och mina kollegor med vetenskapliga metoder visar att detta inte stämmer, det vill säga att läsförmågan inte låter sig mätas riktigt så enkelt som man förutsatt, kort sagt om vi visar upp en svart svan där utbildnings-

systemet tidigare antagit att det bara finns vita svanar, då har vi kanske i alla fall bidragit med om inte *säker* kunskap så i varje fall *säkrare* kunskap, och förhoppningsvis alltså även *meningsfull* och *användbar* kunskap.

Avslutning och framtid

Richard Feynman menade att den högsta dygden av alla för en forskare inte är tvärsäkerhet utan tvärtom undran och osäkerhet, eller kanske skulle vi kalla det *organiserad skepsis* (jfr Merton, 1942). En forskare är aldrig säker på något. Enligt Feynman är det en ovetenskaplig hållning att uttala sig som expert. Vetenskapen lär oss snarare att vi ska förhålla oss skeptiska till expertis, men gravallvarliga i frågan om att undanröja allt det som får oss att tveka på ett påstående som antas vara resultatet av en vetenskaplig studie.

Feynman var briljant. De flesta universitet i världen delar emellertid inte helt hans smala definition av vad vetenskap är. Allt som inte är objektivet och beräkningsbart är inte nödvändigtvis ovetenskapligt. Forskningen om läsning och läsförståelse har kommit ganska långt sedan Henrik II:s dagar, eller sedan införandet av katekesförhören i svenska hus och hem. Vi är dock fortfarande långt från att exempelvis exakt kunna beskriva vilka slags kognitiva och emotionella processer som sätts i spel när ögat möter en rad av små, svarta krumelurer på ett vitt papper eller på en skärm. Vi vet naturligtvis en hel del om dessa processer och har utvecklat ett tekniskt språk för att kunna beskriva dem, men fortfarande är det något ofantligt märkvärdigt och långt ifrån beräkningsbart som sker i det ögonblicket och som låter vårt medvetande kastas från det fysiska rummet och dess begränsningar ut i ett helt annat tillstånd, där läsaren plötsligt får möjlighet att se världen, eller kanske rentav en föreställd, ännu inte existerande värld, med någon annans ögon, ur någon annans perspektiv.

Det är något som jag skulle vilja veta mer om. Även om jag gissar att jag aldrig kommer att kunna beräkna eller helt förutse de processer som tar form när läsare och text möter varandra.

Karlstads universitet har dock uppdragit åt mig att grubbla vidare över dessa saker och även över dess vetenskapliga kvalitet genom att utse mig till professor i pedagogiskt arbete. Det är jag naturligtvis djupt hedrad över. Och jag lovar att försöka få beslutet

Michael Tengberg: *Hängning, tveksamma tester och andra goda motiv för vetenskapliga studier om läsning*

om utnämningen att framstå som rationellt genom att, trots begränsningarna vad gäller predicerbarhet inom mitt forskningsområde, göra mitt bästa för att avslöja ytterligare några svarta svanar i form av otillräckligt underbyggda antaganden om mätning av läsförmåga och därmed förhoppningsvis bidra till säkrare kunskap.

Referenser

- Bet. 2017/18:UbU5 Nationella prov – rättvisa, likvärdiga och digitala. Utbildningsutskottet.
- Feynman, R. C. (2005). *The pleasure of finding things out*. New York: Basic Book.
- Götselius, T. (2013). Från hugg och slag till läslust. I J. Björkman & B Fjæstad (red.) *Läsning. RJs årsbok 2013/2014*. Göteborg/Stockholm: Makadam förlag.
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Marton, F. (1992). Orimligt betygsförslag. I *Dagens Nyheter* 1992-08-18.
- Merton, R. K. (1973) [1942]. The normative structure of science. I R. K. Merton (red.) *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.
- Popper, K. (1963). *Conjectures and refutations: The growth of scientific knowledge*. New York: Routledge & Kegan Paul.
- Rasch, G. (1980). *Probabilistic model for some intelligence and achievement tests*. Chicago, IL: University of Chicago Press.
- Skolverket (2019). *PISA 2018. 15-åringars kunskaper i läsförståelse, matematik och naturvetenskap*. Internationella studier, 487. Stockholm: Skolverket.
- SOU 1992:86 Ett nytt betygssystem: slutbetänkande. Utbildningsdepartementet.
- Tengberg, M. & Skar, G. B. (2017). Hur tillförlitligt är det nationella provet i läsning i åk 9? *Utbildning & Demokrati*, 26(2), 113–137.
- Weir, C. J. (2013). The measurement of reading ability 1913–2012. I C. J. Weir, I. Vidakovic, & E. D. Galaczi (red.). *Measured constructs: A history of Cambridge English language examinations 1913–2012* (ss. 103–179). Cambridge: Cambridge University Press.