

# ATT MÄTA SKRIVFÖRMÅGA

En forskningspresentation om provformat,  
reliabilitet, validitet samt sociala aspekter

Marit Sigurdson

BVM nr 4, 2004



ISSN 1652-7313



# **Att mäta skrivförmåga**

– en forskningspresentation om provformat, reliabilitet, validitet  
samt sociala aspekter

Marit Sigurdson

## **Abstract**

Since 1977, the Swedish Scholastic Aptitude Test (SweSAT) is used for selection to higher education. The test contains five subtests measuring verbal and quantitative abilities. The three subtests constituting the verbal part of the test aims to measure word comprehension and reading abilities in Swedish and English. There have, however, been discussions whether SweSAT also should contain a subtest measuring writing ability. Such an addition, it has been argued, would not only broaden the coverage of the abilities being measured, but would also signal the importance of writing skills throughout the educational system. As a response to this request, this report is a survey of research in the area of measuring writing ability, with focus on questions regarding format, reliability, validity and social aspects. The report also highlights issues for research that are relevant to cover in the event that SweSAT would measure writing abilities.

## Innehållsförteckning

<b>1 Bakgrund och syfte .....</b>	<b>1</b>
<b>2 Mätning av skrivförmåga – en definitionsfråga ..</b>	<b>3</b>
<b>3 Provformat.....</b>	<b>5</b>
<b>3.1 Uppgiftstyper – Direkt eller indirekt mätning .....</b>	<b>5</b>
<b>3.2 Val av ämne.....</b>	<b>8</b>
<b>3.3 Val av genre.....</b>	<b>9</b>
<b>3.4 Provtid .....</b>	<b>9</b>
<b>3.5 Datorisering.....</b>	<b>10</b>
<b>4 Reliabilitet.....</b>	<b>10</b>
<b>5 Validitet.....</b>	<b>12</b>
<b>6 Mätsociologi.....</b>	<b>16</b>
<b>7 Högskoleprovet och mätning av skrivförmåga.</b>	<b>17</b>
<b>7.1 Format – bias och pedagogiska konsekvenser .....</b>	<b>17</b>
<b>7.2 Format men även innehåll.....</b>	<b>20</b>
<b>7.3 Ett nytt delprov ger nya utmaningar .....</b>	<b>22</b>
<b>Litteraturförteckning .....</b>	<b>24</b>
<b>Elektroniska källor .....</b>	<b>26</b>

## 1 Bakgrund och syfte

Högskoleprovet har sedan 1977 använts som urvalsinstrument för högre utbildning i Sverige. Provets primära funktion är att välja ut sökande med förmågor som anses ha betydelse för studieframgång. De förmågor som det nuvarande provet syftar till att mäta kan indelas i en *kvantitativ* och en *verbal* del. I den kvantitativa delen ingår delproven NOG (logiskt tänkande kring matematiska problem) och DTK (diagram, tabeller och kartor), medan den verbala delen består av delproven LÄS (svensk läsförståelse), ORD (svensk ordförståelse) samt ELF (engelsk läsförståelse). Högskoleprovets mätning av den sökandes verbala förmågor är sålunda begränsad till mottagaraspekten i en kommunikationssituation. Mätningar av sändaraspekten, det att i olika språksituationer kunna uttrycka sig i tal och skrift, har med hänvisning till praktiska och ekonomiska svårigheter inte genomförts. Icke desto mindre anses denna förmåga vara en viktig förutsättning för studieframgång, liksom den är avgörande i många andra sammanhang. I synnerhet har mätningar av förmågan att skriva (i denna rapport förkortat till *skrivprov*) efterfrågats som en såväl möjlig som önskvärd komponent vid urval till universitet och högskola. Förväntade fördelar med ett skrivprov lyftes bl.a. fram för ett par år sedan i Högskoleverkets utvärdering av Högskoleprovet.

The addition of a test of writing ability would broaden the skills coverage of the SweSAT [Högskoleprovet, min anm.], signal the importance attached to writing, and would also be likely to increase the relationship between test performance and college success. (Högskoleverket 2002, s. 23.)

Syftet med denna rapport är att belysa ett antal frågor kring mätning av skrivförmåga samt ge några ingångar till den forskning som bedrivits inom detta område. Rapporten behandlar främst frågor kring provformat, reliabilitet, validitet samt sociologiska aspekter på skrivprov. Var och en av dessa aspekter är i sig ämnen för en avhandling, men blir här endast summariskt behandlade. Förhoppningsvis kan rapporten ändå förse utvecklingsarbetet vid Högskoleprovet med en grund för fortsatta diskussioner kring detta mätområde.

Forskningsfältet vad gäller mätning av skrivförmåga med avseende på urval till högre utbildning är vidsträckt. Dock gäller huvudparten av referenserna inom detta fält, och sålunda även i denna rapport, amerikansk forskning och i synnerhet den provverksamhet som bedrivs vid Educational Testing Service (ETS). Vid detta center har bland annat en översikt i ämnet publicerats (Breland, Bridgeman och Fowles, 1999, *Writing Assessment in Admission to Higher Education: Review and Framework*) vilken utgör en stomme till denna rapport.

Att relatera Högskoleprovets verksamhet till den nordamerikanska forskningen innebär givetvis en översättningsproblematik då det svenska utbildningsväsendet, såväl till sin inre pedagogiska som till sin yttre organisatoriska struktur, väsentligt skiljer sig från det amerikanska. De amerikanska antagningsproven till högre utbildning (t.ex. the Scholastic Aptitude Test, SAT, för antagning till college) har heller inte samma funktion eller förutsättningar som Högskoleprovet, vilket ger andra villkor för forskningen. Man bör också beakta skillnader i de socio-kulturella och politiska miljöer där forskning och provverksamhet bedrivs vilket föranleder skilda angreppssätt i frågor om t.ex. sociala grupper och pedagogik. Med detta i åtanke förblir dock många aspekter på mätning och standardiserade prov, diskussioner kring validitet och reliabilitet o.s.v. generaliserbara.

Rapporten är uppbyggd så att det följande kapitlet belyser komplexiteten i begreppet skrivförmåga och i mätningar av denna. I Kapitel 3 redovisas ett antal aspekter på formatet som blir styrande för såväl provkonstruktionsprocessen som det praktiska genomförandet av ett skrivprov. Kapitel 4 och 5 fokuserar på validitet och reliabilitet, främst utifrån de två uppgiftstyperna: essäuppgifter och flervalsuppgifter. Kapitel 6 har fått namnet "Mätsociologi" och vidrör sociala konsekvenser av skrivprov, dess syften och konstruktion. I det sista kapitlet förs en övergripande diskussion kring skrivprov med utgångspunkt i Högskoleprovets förhållanden och med utblick mot framtida forskningsmöjligheter.

## 2 Mätning av skrivförmåga – en definitionsfråga

Att kunna beskriva en händelse, berätta en historia, kritisera ett debattinlägg, föreslå en lösning, revidera en rapport, välja en grammatisk konstruktion o.s.v. är exempel som vi kan härröra till begreppet *skrivförmåga*. Enligt tidigare nämnda studie av Breland m.fl. (1999a) innebär skrivförmåga olika saker i olika situationer och inbegriper en mängd olika kunskaper och kognitiva förmågor. Därigenom blir det också omöjligt att nå en enhetlig definition av vad skrivförmåga är. Men även om man skulle nå en sådan, menar Breland m.fl., skulle inget enskilt prov kunna mäta hela förmågan. (s. 1.)

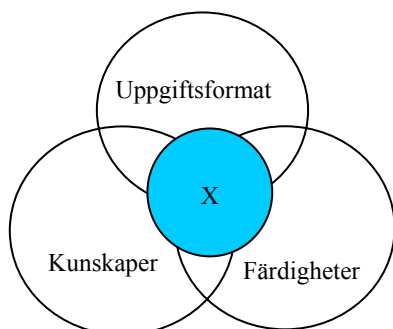
Det finns de som har ifrågasatt huruvida skrivförmåga är mätbart överhuvudtaget. Wiggins skriver 1993 (citerat hos Powers, Fowles och Boyles, 1996, s 6): “Whatever assessors are testing in a twenty-minute essay (such as found in some tests) ... most certainly is not the ability to write.” Vidare har skrivprovets relevans kritiserats då olika utbildningsområden tenderar kräva olika förmågor hos de sökande. På samma sätt har det argumenterats att olika discipliner tillämpar skilda stilar för sin textproduktion, beroende på texternas innehåll och syften. Powers m.fl. (1996) menar dock att mätningar av skrivförmåga har ett visst, om än begränsat syfte.

Perhaps the best that can be hoped for as a result of giving a relatively brief standardized measure is a reasonably clear “snapshot” of an examinee’s skills, rather than a full-length “video”. When properly taken, however, even a simple photograph can often be remarkably revealing of its subject, provided that the photographer’s aim is true and that extraneous influences do not unduly distort the image. (s. 6)

I enlighet med denna metafor blir det viktigt att ställa in siktet och definiera vad man avser att mäta med ett skrivprov. Breland m.fl. (1999a) skriver:

... testing programs need to be very clear about the specific skills and knowledge a test is designed to assess so that test users can determine whether the construct that is actually assessed is appropriate for a particular purpose and population. (s. 1)

Vidare måste man skilja på en eventuell definition av skrivförmåga och den del av förmågan som ett skrivprov faktiskt mäter. Breland m.fl. (1999a) refererar till Willingham och Coles modell av de, delvis sammanfallande, faktorer som påverkar en mätning.



Willingham och Coles modell (Breland m.fl., 1999a)

*Kunskaper* (knowledge) och *färdigheter* (skills) hos provdeltagaren är givetvis faktorer som påverkar, men även *uppgiftsformatet* styr över de kunskaper och färdigheter som kan mätas, även om det i sig inte är av intresse för mätningen. Cirkeln i mitten representerar den del av förmågan som provet mäter. (Breland m.fl., 1999a, s. 2).

De uppgiftstyper som numera ingår i skrivprov, essäuppgifter och flervalsuppgifter (se nedan), inkluderar en mängd möjligheter och begränsningar. Konstruktionen av ett skrivprov måste därför föregås av en gedigen forsknings- och utprövningsprocess som, i samråd med för ändamålet utvalda referensgrupper, definierar vilka aspekter av skrivförmåga som ska mätas. (Breland m.fl., 1999a, s. 2.)

I Willingham och Coles modell blir det tydligt att den förmåga som faktiskt mäts alltid är mindre omfattande än en tänkt, idealisk definition av förmågan. Därför är det även intressant att fråga sig vad som inte mäts. T.ex. mäter skrivprov sällan provdeltagarnas förmåga att planera och revidera sin egen text, trots att studier visar att dessa moment är viktiga för skrivprocessen.<sup>1</sup> (Breland, m.fl., 1999a, s.3.)

---

<sup>1</sup> De praktiska problemen med sådana mätningar, t.ex. låg reliabilitet samt svårigheter med poängbedömningen, har dock ännu inte blivit lösta (Breland m.fl., 1999a, s.3).



### 3 Provformat

Som ovan nämnts styr mätningens format, vare sig man vill eller inte, egenskaperna hos en mätning. Formatet griper över en mängd olika frågor; vilken uppgiftstyp mätningen ska innehålla, hur lång tid provdeltagaren ska ha på sig för uppgifterna, hur provet ska administreras o.s.v. Dessa angelägenheter står givetvis i växelverkan med innehållet i mätningen då t.ex. val av uppgiftstyp beror på vad uppgiften ska bestå av. Detta kapitel kommer att presentera några av de aspekter som kan relateras till ett skrivprovs format.

#### 3.1 Uppgiftstyper – Direkt eller indirekt mätning

Det har traditionellt sett funnits två typer av metoder att mäta skrivförmåga: *indirekt*, där provdeltagaren svarar på flervalsuppgifter relaterade till skrivförmåga och *direkt*, där provdeltagaren de facto skriver något (Powers m.fl., 1996, s.6).<sup>2</sup> För flervalsuppgifter gäller att de vanligtvis mäter provdeltagarens förmåga att förbättra en text genom att välja ett bland flera alternativa uttryckssätt för att nyansera språket, korrigera grammatiken, förbättra meningsbyggnaden, omdisponera textstycken o.s.v. De uppgifter där provdeltagaren skriver något kräver däremot en integrerad användning av dessa förmågor samtidigt som han/hon visar sin förmåga att effektivt frammana, syntetisera och formulera sina egna tankar om ett ämne, ett argument, ett förslag etc. (Breland m.fl., 1999a, s.4.) De direkta mätningarna kallas i engelsk litteratur för *essays* och inbegriper en mängd olika genrer.<sup>3</sup> I denna rapport översätts de direkta mätningarna till *essäuppgifter*, vilket ska tolkas i sin vidaste betydelse som uppgifter som kräver textproduktion, oavsett genre.

---

<sup>2</sup> Dessa kan, i andra sammanhang och på ett högre abstraktionsplan, betraktas som två metoder för indirekta mätningar, där indirekt syftar på att de egenskaper som mäts (i det här fallet kunskaper och färdigheter) är dolda och måste mätas via indikatorer för egenskapen. (Se bl.a. Wolming, 2000, s 83.) Vad gäller skrivförmåga (och såsom begreppen används här) kan direkt mätning även kallas *performance assessment* (se bl.a. Breland m.fl., 1999a) och indirekt mätning *objective assessment* (se bl.a. Cooper, 1984).

<sup>3</sup> Jämför svenskans *essä* som närmast definieras som en genre i sig, om än en vid sådan. Nationalencyklopedin översätter ordet: ”konstnärligt och personligt utformad uppsats eller artikel vanl. över ngt fil., hist. el. estetiskt ämne”.

De två metodernas för- och nackdelar har diskuterats länge, vilket Cooper (1984) redogör för i *The Assessment of Writing Ability: A Review of Research*, 1984. I början av 1900-talet utgick man från att skrivförmåga enbart kan mätas direkt. Men efter att poängbedömningen av essäerna visade sig vara både inkonsekvent utförd och ha låg reliabilitet började utbildningspsykologer på 1920- och 1930-talen att experimentera med indirekta mätmetoder. Dessa metoder visade sig resultera i mycket hög statistisk reliabilitet, men också minskade kostnader i fråga om administration och rättning av det ökande antalet prov. Efter ca 60 års debatt tycktes fördelarna med de indirekta metoderna för storskaliga mätningar av skrivförmåga vara svåröverträffade. Cooper (1984) citerar Orville Palmers konstaterande från mitten av 1960-talet: "essay tests are neither reliable nor valid, and that, whatever their faults, objective tests do constitute a reliable and valid method of ascertaining student compositional ability." (s.1f.)

Dock dröjde det inte länge förrän studier visade på både valida och reliabla essäuppgifter och den direkta mätmetoden vann, enligt Cooper, med tiden mark på flervalssuppgifternas bekostnad. Dels hade de indirekta mätningarna svårt att ernå "face validity" (ung. legitimitet) hos engelsklärarna, dels fanns en oro över att dessa mätningar sände oönskade signaler till elever och lärare: att skrivförmåga inte var viktigt och/eller att det kunde läras in eller läras ut genom flervalssuppgifter. Förutom detta kritiserades de indirekta metoderna för att de inte mäter skrivförmåga på en högre nivå, t.ex. originalitet, kreativitet, retoriska förmågor som stilistisk säkerhet, målgruppsanpassning o.s.v. Vid skrivande stund för Cooper (1984) bestod enbart 5 % av alla landsomfattande prov av uteslutande flervalssuppgifter. (s. 2f.)

Förutom mätteoretiska och pedagogiska hänsynstaganden är givetvis konstruktionen av standardiserade prov även underlagda ekonomiska villkor, vilket får konsekvenser för valet av uppgiftstyp. Breland m.fl. (1999a) skriver:

The challenge for writing committees is to design an assessment that measures an appropriate range of writing skills and yet is also cost-effective and efficient. Rarely does a testing program administer a single writing task without a multiple-choice section or additional writing tasks. (s.4.)

Detta påstående stämmer om man ser vilka uppgiftstyper som ingår i de nationellt administrerade prov, här benämnt provprogram, som utvecklas vid olika amerikanska center. En kombination av uppgiftstyperna återfinns bl.a. i *the College Board's SAT program*. I *SAT II: Subject writing test*<sup>4</sup> skriver provdeltagarna en essä på 20 minuter och svarar på 60 flervalsuppgifter (av typen ”känna igen språkfel”, ”förbättra meningar” och ”disposition”) på 40 minuter. Det generella provet *SAT Reasoning test* kommer fr.o.m. våren 2005 att ersättas av *The New SAT*. Den största förändringen från det nuvarande provet blir tillägget av ett skrivprov.<sup>5</sup> Detta delprov innebär en flervalsuppgiftsdel på 35 minuter samt en essä på 25 minuter. I och med denna förändring kommer *SAT II: Subject writing test* att utgå, då detta prov till stora delar liknar *The New SAT*.<sup>6</sup> (Se bl.a. Lawrence, Rigol, Van Essen, Jackson (2003) och [www.collegeboard.com](http://www.collegeboard.com).)

I de provprogram där flervalsuppgifter inte inkluderas består mätningen av två essäuppgifter. De tre provprogrammen *MCAT* (medical colleges), *GMAT* (graduate schools of business management) och *GRE* (graduate schools, all disciplines) har två analytiska skrivuppgifter (på 30 - 45 minuter vardera). De två sistnämnda har dock, utöver skrivproven, en verbal del med flervalsuppgifter (bl.a. ”förbättra meningar” och ordkunskap). Ett undantag från påståendet i citatet ovan är *LSAT* (antagningsprov till ”Law school”) som bara består av en skriftlig argumentation på 30 minuter. ([www.kaptest.com](http://www.kaptest.com), [www.gmac.com](http://www.gmac.com), [www.gre.org](http://www.gre.org) och [www.lsat-center.com](http://www.lsat-center.com).)

I skenet av den historiska översikt som Cooper (1984) ger tycks alltså den direkta mätningen ha överlevt både reliabilitetsproblem och negativa kostnadsaspekter då man i vissa fall enbart använder sig av denna metod. Skrivprov med enbart indirekta mätmetoder, å andra sidan, förekommer inte i något av de stora provprogram som här, i likhet med hos Breland m.fl. (1999a), tagits upp. Den direkta mätningens status i USA återspeglas i antalet forskningsrapporter som

---

<sup>4</sup> *SAT II: Subject tests* efterfrågas av många colleges som ett komplement till betyg, poäng på *SAT Reasoning test* m.m.

<sup>5</sup> *SAT Reasoning test* har i nuläget ingen skrivuppgift utan består bl.a. av en verbal del på 75 minuter som fokuserar på läsning men som även inkluderar "analogies and sentence completions, which emphasize logical relationship, vocabulary, and how words relate". ([www.collegeboard.com](http://www.collegeboard.com).)

<sup>6</sup> *The New SAT:s* flervalsuppgifter kommer att vara av typen ”förbättra meningar” och ”känna igen språkfel”.

publicerats i detta ämne, medan forskning om flervalssuppgifternas egenskaper för mätning av skrivförmåga är betydligt mer sällsynt. Därför blir följande presentation av ytterligare några aspekter på formatet främst begränsad till direkta mätningar av skrivförmåga.

### **3.2 Val av ämne**

För standardiserade prov gäller generellt ambitionen att alla provdeltagare ska ställas inför samma krav. I ett flervalssuppgiftsprov försöker man oftast uppnå detta genom att ge alla samma uppgifter. Det är givetvis möjligt att även för en essäuppgift ge alla provdeltagare ett och samma ämne, men om syftet är att mäta förmågan att uttrycka sig i skrift, och inte förmågan att uttrycka sig om ett visst ämne t.ex en historisk händelse, blir en sådan mätning både orättvis och får låg validitet. Breland m.fl. (1999a) skriver: "Allowing some degree of choice within a single question [...] would improve standardization and validity for a test of writing skills". Detta mått av valbarhet kan uppnås genom att man antingen ger ett ämne som tillåter stora variationer av svar, eller att man ger möjlighet att välja mellan flera olika ämnen. (1999a, s.6f.)

Att ge valfrihet bland flera ämnen medför dock vissa utmaningar. Bl.a. måste alla ämnen vara likvärdigt svåra så att inte provdeltagaren blir otillbörligt straffad av ett felaktigt val. Likaså måste man framställa jämbördiga bedömningsskalor för rättning. Vidare kan själva valet för provdeltagaren innebära en tidsförlust som påverkar hans eller hennes förmåga att färdigställa uppgiften. (Breland m.fl., 1999a, s. 7.)

Eftersom essäämnen tenderar att vara kortfattat formulerade och därför lätta att minnas är det också svårt att säkerställa att ingen provdeltagare får tag på ämnena innan provet och därmed vinner fördelar. Datorbaserade prov som *GMAT* och *GRE* offentliggör i förväg en samling av essäämnen, däribland de som kommer på provet. Man utgår då från att om de offentliga ämnena är tillräckligt många (ca 70-150) så blir det inte rimligt för provdeltagaren att förbereda svar till alla ämnen. (Breland m.fl. 1999a, s. 7.)

### 3.3 Val av genre

Precis som provdeltagare är olika rustade för att skriva om vilket ämne som helst är de också olika rustade för att skriva i vilken texttyp eller genre som helst, skriver Cooper (1984). Olika genrer (t.ex. beskrivande, berättande, utredande eller argumenderande) ställer olika krav på skribenten vilket påverkar provresultaten. Cooper presenterar ståndpunkten att: ”The implication is that writing for different aims draws on different skill constructs which must therefore be measured and reported separately to avoid erroneous, invalid interpretations of performance, as well as inaccurate decisions based on such performances.” (s. 5.)

### 3.4 Provtid

Tilldelningen av tid för ett prov måste relateras till syftet med provet, förmågan man vill mäta, populationen (d.v.s. deltagarna), användandet av provpoängen etc. Om man ser till de amerikanska provprogrammen i redovisningen ovan varierar provtiden för essäskrivning mellan 20-45 minuter. T.ex tilldelas provdeltagaren bara 20 minuter för essädelen i *SAT II: Writing Subject Test*, vilket av Breland m.fl. (1999a) förklaras med att den främst baseras på personliga erfarenheter och inte kräver någon analys. När man istället vill mäta en mer analytisk förmåga, som i *GRE* och *California State Universitys English Placement Test*, tillåts 45 minuter per essä. (s.5.)

Ju längre tid som tillåts för en skrivuppgift desto högre poäng tenderar provdeltagaren att få, menar Breland m.fl. (1999a) och visar på studier av essäuppgifter i bl.a. *TOEFL*, *SAT II: Writing Subject Test* och *GRE*. Det behöver för den delen inte betyda att provtiden påverkar hur provdeltagarna rankas. Resultatet från *TOEFL*-studien visar att rankingen är oförändrad trots en förlängning av provtiden från 30 till 45 minuter. Dock är resultaten inte entydiga vad gäller detta. För *SAT II* visas att en förkortad provtid gynnar kvinnor, medan åter en annan studie av *California Bar Examination* inte ger något samband mellan kön och provtid. Vidare tycks inte en kortare provtid missgynna provdeltagare som identifierar sig som långsamma skribenter, enligt studien av *GRE*. Samma studie konkluderar att olika provtider inte ger någon märkbar effekt på korrelationer mellan essäpoäng och andra

kriterier för skrivförmågan. Breland m.fl. konkluderar att provtiden måste baseras på rikliga utprövningsdata. Först när man analyserat resultatet från representativa grupper skrivprestationer, under riktiga provsituationer, kan man rekommendera en provtid för en viss uppgift. (s.5.)

### **3.5 Datorisering**

En fråga som blir alltmer aktuell - och vars svar sätter villkor för hela provverksamheten från konstruktion till administration - är huruvida provet ska vara datoriserat. Redan 1999, när Breland m.fl. (1999a) publicerade sin översikt om skrivprov, blev flera storskaliga provprogram i USA (bl.a. *GMAT* och *TOEFL*) administrerade via dator och datoriseringen tycks öka i takt med att utvecklingen går framåt. Bl.a. har rättningen av essäuppgifter alltid inneburit en enorm kostnad i både tid och pengar, då ett stort antal läsare samlats under ett visst antal dagar för att, efter så likvärdiga kriterier som möjligt, rätta provdeltagarnas texter. Numera finns det exempel på program där läsaren från sin hemdator, och när som helst, kan utföra sitt rättningssupdrag. Programmet är också utvecklat för att kontinuerligt kontrollera kvaliteten på läsarens förmåga att rätta. (Se bl.a. e-rater på [www.ets.org](http://www.ets.org)). Så här skriver Högskoleverket, 2002, om utvecklingen inom rättning:

[ T]he task of scoring a large number of essays would be a formidable one, but this is the kind of step taken by a number of programs in the US. There are ways, moreover, of applying technology to facilitate the grading enterprise. Choices range from using the Internet to distribute essays to readers in remote sites to actually employing natural language processing systems to grade student papers. The idea of computers grading essays, once viewed as quite futuristic, has turned around dramatically in the US, so that one can now say that the future is now!" (s. 23.)

## **4 Reliabilitet**

En angelägenhet för varje provutvecklare är mätningens reliabilitet, d.v.s. egenskapen att vid upprepade tillfällen ge samma resultat, förutsatt att omständigheterna är desamma. Det finns en mängd

faktorer som påverkar denna säkerhet i mätningen. Breland m.fl. (1999a) poängterar två:

Essay test reliability is influenced by at least two major sources of measurement error: (1) the degree to which the assessment domain is sampled, and (2) rater disagreement. (s. 25.)

Angående det förstnämnda, i vilken utsträckning mätningar görs, gäller generellt att reliabiliteten ökar med antalet uppgifter i ett prov. Som nämndes i kap 3.1 har indirekta mätningar med flervalstuppgifter också en högre reliabilitet än essäuppgifter, där antalet uppgifter av praktiska skäl är färre. Brent Bridgeman (1991) skriver i en artikel i *Research in Higher Education*:

... because of measurement error created by subjective scoring, and by relatively narrow coverage of the content domain (there are usually only one to three essays on one test), essay tests are usually substantially less reliable than multiple-choice tests in the same general subject area." (s 320.)

I samma anda skriver Breland m.fl. (1999a): "A single essay test is not likely to yield individual scores that are reliable enough for admission purposes" (s.4). Likaså konstaterar Breland, Kubota, Nickerson, Trapani och Walker. (2004) i en undersökning av gruppskillnader och reliabilitet vad gäller *the New SAT* att: "The reliability estimates obtained in the current study, as well as those obtained in previous studies, indicate that essay assessments of writing skill are not as reliable as most traditional educational assessments" (s. 9).

Även om en enda essä ger för låg reliabilitet är det å andra sidan sällan praktiskt eller ekonomiskt möjligt för storskaliga mätningar att administrera fler än två essäer. Därför blir det intressant att undersöka reliabiliteten hos mätningar med två essäer, menar Breland.m.fl. (1999a) och presenterar sex olika studier gjorda mellan 1951 och 1987. Dessa mätningar visar i snitt en reliabilitet på .71. I en senare studie av reliabiliteten i *MCAT*:s skrivprov (vilket består av två essäer som rättas av två oberoende läsare) uppmäts reliabiliteten med split-half –metoden till .77 - .80 och med test-retest-metoden till .61 - .68. (1999a, s. 13.)

Reliabiliteten med direkta mätningar skulle givetvis öka om bedömningen baserades på flera essäer om olika ämnen, skrivna på

olika dagar och rättade av olika läsare. Vad gäller det sistnämnda hänvisar dock Breland, m.fl. (1999a) till en översikt av reliabilitetsstudier där antalet läsare, utöver tre, inte visade sig ha någon större effekt på reliabiliteten medan det för antalet uppgifter gällde att ju fler, desto högre reliabilitet. Vidare skriver Breland m.fl: ”This review, as well as individual studies, indicates that what is most important in maximizing the reliability of essay tests is the number of tasks used.” (s.13.)

Slutsatsen hos Breland m.fl. är flervalsuppgifter eller flervalsuppgifter i kombination med essäuppgifter ger den högsta reliabiliteten vad gäller skrivprov, med värden mellan .80 - .90. T.ex. har *SAT II Writing Subject Test* (bestående av 40 flervalsuppgifter på 60 minuter och en essä på 20 minuter) över flera provtillfällen visat en reliabilitet på .88 till .91 för flervalsuppgiftsdelen och .85 till .91 för provets sammanlagda poäng (1999a, s 14 och 25).

## 5 Validitet

”It is better to have a somewhat inconsistent measure of the right thing than a very consistent measure of the wrong thing”, resonerar Cooper (1984, s. 9) och menar därmed att validiteten är viktigare än reliabiliteten. Man bör dock klargöra vad man menar med *validitet* då begreppet kan avse olika aspekter på mätningen. Här kommer främst begreppsvaliditet och prediktiv validitet hos skrivprov att diskuteras.

Vad gäller begreppsvaliditeten påpekar Cooper faran med att förväxla denna med *face validity* (s 3). I fallet med skrivprov har, som tidigare nämnts, direkta mätningar haft högre *face validity* än indirekta eftersom de kräver att provdeltagaren utför just det som ska mätas. Men som Cooper skriver: ”the validity of direct assessment and the invalidity of indirect assessment cannot simply be presumed because one requires writing and the other does not” (s 10). I enlighet med detta menar Powers, Fowles och Willard (1993): ”the apparent directness of such [direct] measures does not exempt them from the need to marshal validity evidence in support of their use” (abstract).



Vad gäller skrivprovets *begreppsvaliditet*, d.v.s. huruvida mätningen kan mäta det man definierat som skrivförmåga, diskuteras detta främst hos Cooper (1984), som presenterar kritiska åsikter vad gäller de direkta mätningarna:

If an indirect assessment can be said to measure mostly editorial skills and error recognition, a brief direct assessment can be said to measure the ability to dash off a first draft, under an artificial time constraint, on a prescribed or hastily invented topic that the writer may have no real interest in discussing with anyone but must discuss with a fictional audience. (s. 9.)

Citatet summerar de invändningar som Cooper förmedlar i sin forskningsöversikt. Han refererar till påståenden om att tiden för direkta mätningar ofta är alldeles för kort, vilket gör att mätningen inte säger något om vad provdeltagaren kan prestera på längre tid. En kort provtid strider även mot en pedagogisk syn på skrivandet som en process, vilken kräver tid för eftertanke och revideringar. Vad gäller ämnesvalet poängterar Cooper svårigheten att hitta ämnen som inte är för specifika och kontextbärande då de därmed påverkar olika beroende på person och tidpunkt. Ämnena får heller inte vara för generella, menar han, eftersom de då väcker ett varierat engagemang från provdeltagarna, vilket också sänker reliabiliteten. (s. 4 och 6.)

En annan kritik är det artificiella sammanhang där den direkta mätningen genomförs. Ofta anges inte någon målgrupp, skriver Cooper, och även om så görs beror resultatet på provdeltagarens föreställningsförmåga. Vidare innebär en specificerad målgrupp att provdeltagarna skriver för två målgrupper: en fiktiv och en verklig, varav den senare ska ignoreras. ”At best, scores will depend on how well candidates recognize and apply the conventions of standard middle-class usage”, skriver Cooper. (s. 9 och 11.)

De indirekta mätningarna å andra sidan, menar Cooper, har större begränsningar när det gäller vad de kan mäta, men medger också större kontroll över det man mäter. Flervalssuppgifter kan fokusera på specifika aspekter av skrivförmåga medan bedömningen av en provdeltagares prestation på en essäuppgift kan grundas på en mängd aspekter, d.v.s. den kan underkännas eller godkännas p.g.a. grammatiska felaktigheter och stavfel såväl som på grund av bristande logik eller disposition. (s 3.)

Cooper menar också att de indirekta mätningarna kan mäta skrivförmåga och rangordna provdeltagare i samma utsträckning som erfarna läsare kan bedöma en skribents förmåga efter att ha läst en stor mängd av dennes textproduktion. Han presenterar ett antal studier där man funnit en hög korrelation mellan indirekta och direkta mätningar. Bl.a. presenteras Godshalks studie från 1966 vars syfte var att kontrollera validiteten hos flervalssuppgifterna i *ECT* (föregångaren till *SAT II: Subject writing test*). För att göra denna kontroll användes som kriterium fem essäuppgifter, rättade av fem läsare. Korrelationen mellan flervalssuppgifterna och den totala poängen på essäerna visade sig vara högre (.75) än vad korrelationen i snitt var mellan de separata essäerna (.52). (s. 12f.)

I en mer nylig studie undersökte Power m.fl. sambandet mellan poäng på essäuppgifter och andra kriterier för skrivförmåga. Det starkaste sambandet fann man med kursrelaterade skrivuppgifter och det lägsta med studenternas bedömning av deras egna prestationer. (1999, s. 13.)<sup>7</sup>

Senare validitetsstudier har annars främst gällt *prediktiv validitet*, d.v.s. i vilken utsträckning mätningen kan förutsäga en provdeltagares framtida prestationer.<sup>8</sup> Breland m.fl. (1999a) presenterar undersökningar av skrivprovs prediktionsförmåga utifrån dess korrelation med tre olika typer av kriterier: (1) betyg och medelbetyg<sup>9</sup> eller lärares bedömningar, (2) skrivprestationer, (3) olika faktorer i prov-deltagarnas bakgrund. Skrivprestationer har i två undersökningar använts som kriterium för samtliga varianter av provtyper (en essä/ två essäer/ flervalssuppgifter/ flervalssuppgifter + essä). Enligt dessa studier var medelvärdet av korrelationerna för den prediktiva validiteten lägst vad gällde prov med en essä (.49) och högst hos prov där

---

<sup>7</sup> Undersökningen gällde relationen mellan två typer av essäuppgifter ("Issue" och "Argument") i *GRE* och ett antal kriterier på den skrivförmåga som kan krävas i akademiska sammanhang.

<sup>8</sup> Även här får man påminna sig om att prestationer kan bedömas olika i USA och i Sverige, liksom att innebörden av ett framgångsrikt skrivande kan skilja sig mellan olika discipliner (se bl.a. Cooper, s.23).

<sup>9</sup> USA har, till skillnad från Sverige, ett tämligen logiskt och definierat kriterium i det första collegeårets betyg (FGPA). Dock har det amerikanska betygssystemet, i jämförelse med det svenska, en mycket lokal tillämpning vilket försämrar jämförbarheten mellan skolor. (Se bl.a. Skolverkets remissvar till SOU 2004:29, *Mellan myt och verklighet*, s 49 f.)

flervalsuppgifter och en essä kombineras (.73). Prov med enbart två essäer och prov med enbart flervalsuppgifter erhöll ungefär samma korrelationer (.67 respektive .66). (s. 14ff.)

Liknande resultat fann Breland, Kubota och Bonner (1999b) när de undersökte *SAT II: Writing Test:s* prediktionsförmåga vad gäller skrivprestationer på college. Utifrån tre kriterier, liknande de ovan nämnda,<sup>10</sup> fick provdelen med flervalsuppgifter (40 minuter) högre prediktionsförmåga än den 20 minuter långa essäuppgiften. Även här återfanns den högsta prediktiva validiteten i kombinationen av de båda uppgiftstyperna. (s. 10.)

I den nyligen publicerade studien av *the New SAT*, av Breland, m.fl. (2004) framkom samma skillnader mellan uppgiftstyperna vad gäller korrelationer för den prediktiva validiteten, även om vissa av dessa korrelationer inte var lika höga som i andra studier, bl.a. Godshalks som presenterades av Cooper (1984). Detta kan dock delvis förklaras av att tidpunkten för insamlandet av data för kriteriet och prediktorn sammanfallit, i Godshalks studie, medan man här samlat in data före och efter antagningen till college. Vidare tillät man här ett val bland flera ämnen för essän, vilket komplicerade rättningen och sänkte reliabiliteten, menar författarna. (2004, s. 10.)

En annan angreppsvinkel, men parallella resultat visar Bent Bridgeman (1991). Där behandlas i vilken utsträckning de två uppgiftstyperna kan prognosticera studieframgång i college<sup>11</sup> utöver vad medelbetyg från high school och *SAT* gör. Resultaten visade att flervalsuppgifterna (av typen ”känna igen skrivfel”) bidrog till den prediktiva validiteten med en korrelation på .01-.02 medan essäuppgiften praktiskt taget inte tillförde någonting till dessa poäng. (s. 330.)

---

<sup>10</sup> Som kriterier användes provdeltagarens (1) betyg i engelska, (2) skrivprestation på fyra hemuppgifter, bedömd av 20 erfarna läsare samt (3) bedömning av den egna skrivförmågan (s. 9).

<sup>11</sup> Prediktionsförmågan definierades alltså här utifrån *medelbetyget* efter första året på college (Freshman Grade Point Average) och inte specifikt utifrån skrivprestationer på college.

## 6 Mätsociologi

Det nykonstruerade ordet, mätsociologi, i rubriken hänvisar till mätningens roll i samhället samt de sociala frågor som kringgärdar denna roll. Konkret sett har antagningsprov en uppenbar effekt på representationen av grupper inom universitet och högskola. I detta kapitel kommer skrivprovets mätsociologi att diskuteras dels utifrån noterade prestationsskillnader mellan grupper av provdeltagare, dels utifrån de effekter som dessa skillnader kan leda till.

Breland m.fl. (1999a) påpekar att skillnader mellan gruppers prestationer på skrivprov nästan alltid är mindre på prov med låg reliabilitet. Ju högre andel ”störningar” i poängen - vilka fördelas slumpmässigt och därmed jämnt över populationen - desto mindre skillnader i resultaten. Om ett essäprov är mindre reliabelt än ett flervalsprov uppvisar det också mindre skillnader mellan olika grupper. Vidare beror både gruppskillnaders storlek och riktning på vad det är man mäter. Sätillvida skulle ett prov med flervalsuppgifter och ett prov med essäuppgifter, utifrån antagandet att de två uppgiftstyperna mäter olika delar av skrivförmåga, resultera i olika gruppskillnader även om de var likvärdigt reliabla. (s. 20.)

Med dessa reservationer i åtanke finns det dock ett antal studier som kan belysa gruppskillnader vad gäller prestationer på skrivprov. T.ex. konkluderar Breland m.fl. (1999a) att könsskillnaderna generellt är små vad gäller båda uppgiftstyperna, men att kvinnor relativt sett presterar bättre än männen på essäuppgifter än på flervalsuppgifter. Männen presterar däremot bättre än kvinnorna på de verbala prov som är obligatoriska vid många provprogram. Införandet av essäuppgifter i en mätning innebär alltså, enligt författarna, att en högre andel kvinnor antas till högre studier. (1999a, s. 26.) I Breland m.fl. (2004) framkommer liknande resultat om *the New SAT*: ”The results of this study, as well as those of previous studies, indicate that female students can be expected to score significantly higher than male students on an essay assessment of writing skill” (s. 9). I studien påpekas dock att denna skillnad till kvinnornas fördel varierar i storlek beroende på population (s 1).

I studier där man jämfört den vita populationen med andra etniska grupper (indelad på provdeltagare med asiatiskt, afrikanskt och

latinamerikanskt ursprung) vad gäller prestation på essäuppgifter har man funnit anmärkningsvärda skillnader. Beroende på grupp och utbildningsnivå varierar denna skillnad från .33 - .75 standardavvikelseenheter. (Breland m.fl. 2004, s.1.)

Skillnaderna mellan den vita populationen och asiatiska och spanska grupper som inte har engelska som sitt modersmål är dock mindre för essäuppgifter än för flervalsuppgifter, visar Breland m.fl. (1999a). Detta kan, enligt författarna, förklaras dels med att essäuppgifter tillåter provdeltagaren att undvika obekanta formuleringar och ord medan flervalsuppgifter ofta kräver en igenkänning av dessa, men också med att essän bedöms utifrån många andra aspekter än grammatik och syntax (bl.a. disposition och idéer), vilket fokuseras i flervalsuppgiftsprov. (s 21.)

## **7 Högskoleprovet och mätning av skrivförmåga**

I denna rapport framträder en del av den forskning som bedrivits kring några aspekter av mätning av skrivförmåga. Rapporten ger alltså ingalunda en heltäckande översikt av forskningsfältet. Denna rapport speglar också, med sin fokus på amerikansk forskning, bristen på studier med utgångspunkt i svenska förhållanden i allmänhet och antagning till högre utbildning i synnerhet. Önskemål om att införa ett skrivprov i Högskoleprovet möter därmed en mängd frågor och utforskade problemområden. Med tanke på det sistnämnda finns det anledning att i detta avslutande kapitel knyta diskussionen till Högskoleprovets verksamhet och utveckling.

### **7.1 Format – bias och pedagogiska konsekvenser**

Förutom mer övergripande principiella frågor kring hur Högskoleprovet kan mäta skrivförmåga, är en konkret fråga hur ett sådant prov skulle se ut. Stora delar av rapporten har just också behandlat aspekter utifrån skrivprovets format, då detta ofrånkomligen ger konsekvenser för mätningen. Breland m.fl. (1999a), lyfter fram två konsekvenser av formatet som man måste ta i beaktande. Det ena, vilket kan kopplas till föregående kapitel, är bias d.v.s. systematiska över- eller underskattningar av grupper:

The first relates to bias. If it can be shown that one writing test format favors some groups of examinees over others for reasons unrelated to the construct, then that is evidence of bias. (s. 10)

För det nuvarande Högskoleprovet gäller att olika sociala grupper presterar olika bra.<sup>12</sup> Bl.a. har Reuterberg (2003) visat på betydelsen av utländsk bakgrund för resultat på provet. De provdeltagare som är födda i Sverige och har minst en svenskfödd förälder får de högsta resultaten medan de som har invandrat (spec. efter 10 års ålder) och har minst en förälder född i utlandet får de lägsta resultaten. Det prov som personer med utländsk bakgrund klarar bäst är ELF, vilket tyder på att bristande kunskaper i svenska är av betydelse. (s.31ff.)

I enlighet med Breland m.fl. (1999a, se kap. 6) finns det anledning att tro att de som har svenska som andraspråk klarar ett essäprov bättre än vad de klarar flervalsuppgifter, relativt provdeltagare med svenska som modersmål. Utifrån detta skulle man kunna anta att tillägget av ett essäprov gynnar provdeltagare med utländsk bakgrund, vilket skulle ligga i linje med strävan att öka rekryteringen av sådana studenter. Dock utgår resonemanget från att provdeltagarna kan undvika svårigheter och prestera bra på en essäuppgift trots bristande skrivförmåga, vilket väcker frågor kring validitet och vad som är skrivförmåga.

Högskoleprovet har också en könsrelaterad prestationsskillnad, där kvinnor i genomsnitt får ett lägre resultat än män. Denna skillnad kan delvis förklaras av att provet, p.g.a. det frivilliga deltagandet, är utsatt för självselektion.<sup>13</sup> Dock kan införandet av ett skrivprov ge konsekvenser för könsfördelningen i urvalsgruppen då kvinnor, enligt de amerikanska studierna tenderar prestera bättre på skrivprov än män, oavsett uppgiftstyp.

---

<sup>12</sup> Att prestationsskillnaderna mellan grupper blir tämligen synliga kan förklaras av att Högskoleprovet har en hög reliabilitet, ca. 0,92 enligt KR20 (se kap. 5).

<sup>13</sup> Bl.a. deltar fler kvinnor än män i provet medan de män som deltar i snitt har högre utbildning och kvinnorna i snitt är äldre. Selektionen tycks också vara större bland männen än bland kvinnorna. Vad gäller prestationsskillnaderna har dock gymnasieutbildning visat sig ha betydligt större betydelse än kön. Se bl.a. Stage och Ögren (2003).

Som antydde i kap. 6 bör prestationsskillnader mellan olika grupper beaktas utifrån de politiska konsekvenser som de medför. Samtidigt kan de amerikanska forskningsresultaten ingalunda utan vidare överföras till de svenska populationerna. Det krävs därför ett gediget utprövningsförfarande under svenska förhållanden för att konstruera ett prov som mäter olika sociala gruppers skrivförmåga på ett acceptabelt sätt. Frågor som då blir relevanta att besvara är bl.a. vilka effekter de båda uppgiftsformaten har på olika gruppers prestationer. Bidrar t.ex. ett essäprov till högre prestationer bland provdeltagare med utländsk bakgrund?

Den andra konsekvensen som Breland m.fl. (1999a) belyser är eventuella styreffekter på den pedagogiska verksamheten som provet kan ge:

The second relates to educational consequences. If the writing test format influences curriculum because teachers 'teach to the test' or because they or their students place undue emphasis on question types found on national tests, then the test can be viewed as having an influence on the educational system. If people agree that this influence is positive, that is all to the good. The problems arise when at least some groups claim that the format has a negative influence on instruction and learning. (1999a, s. 10.)

Sedan mitten på 1970-talet har den amerikanska forskningen engagerats kring frågan om provens styreffekter på utbildning. Vad gäller skrivprov har man, som nämndes i kap. 3.1, i en strävan mot mer autentiska mätningar argumenterat för essäuppgifter, då flervalsuppgifter ansetts bidra med oavsiktliga sidoeffekter i lärares och elevers syn på språkinläring. Dock har man inte, enligt Breland m.fl. (1999a), empiriskt undersökt huruvida standardiserade prov för urval till högre utbildning faktiskt påverkar undervisningen. (s. 10.) Icke desto mindre framförs de pedagogiska konsekvenserna som ett starkt skäl bl.a. till att man nu inför ett prov på skrivförmåga i *SAT*.

The new SAT will improve the alignment of the test with current curriculum and institutional practices in high school and college. By including a third measure of skills -- writing, the new SAT will help colleges make better admissions and placement decisions. In that way, the new SAT will reinforce the importance of writing throughout a student's education. ([www.collegeboard.com](http://www.collegeboard.com))

Det är inte känt huruvida Högskoleprovets innehåll och format påverkar t.ex. gymnasieskolans undervisning och i så fall hur dessa styreffekter ser ut. Ändå framförs argumentet, bl.a. av Högskoleverket, att ett skrivprov i Högskoleprovet skulle förmedla vikten av skrivförmåga till landets skolor (se kap. 1). De pedagogiska konsekvenserna avgörs dock inte enbart av *att* skrivförmåga mäts utan också av *hur* den mäts. T.ex. måste man fastställa vilka kunskaps- och färdighetsområden (enligt t.ex. Williams och Coles modell i kap. 2) som skulle kunna mätas i ett skrivprov av Högskoleprovets omfattning och vilken påverkan detta skulle ha på t.ex. svenskämnet i gymnasieskolan. Man får då bl.a. fråga sig, i enlighet med problematiken som belystes i kap. 5, hur ett standardiserat skrivprov med tusentals provdeltagare kan svara mot gymnasieskolans målsättning om ett mer processororienterat arbetssätt. De pedagogiska konsekvenserna handlar därmed i lika hög grad om ett skrivprovs validitet och legitimitet, samt hur väl detta förankras hos allmänheten. Om provet inte lyckas möta dessa krav riskerar styreffekterna istället att bli negativa.

## **7.2 Format men även innehåll**

Vad gäller uppgiftsformatet förordas i studierna en kombination av indirekt och direkt mätning av skrivförmåga då formaten kompenserar för varandras brister. Detta gäller både med avseende på reliabiliteten och för validiteten under amerikanska förhållanden. För att kunna bekräfta eller förneka dessa relationer för svensk del krävs åtskilliga studier av de två uppgiftsformaten utifrån den population som genomför Högskoleprovet. T.ex. kan man fråga sig: Hur korresponderar de två uppgiftstyperna med uppfattningar om och definitioner av skrivförmåga hos provets avnämare? Vilken relation har provdeltagarnas resultat på de olika uppgiftstyperna med deras gymnasiebetyg, resultat på andra verbala prov o.s.v.? Vad skulle de respektive uppgiftstyperna bidra med, mätteoretiskt sett, till det befintliga provet? Vad är t.ex. sambandet mellan en essäuppgift och ORD-delen på Högskoleprovet?

De amerikanska studier som uppmärksammats i denna rapport belyser en mängd mättekniska skillnader mellan indirekta och direkta mätning. Dock analyseras inte skrivprovets innehåll på något utförligt



sätt i dessa studier, med undantag för Cooper som visar på uppsatsämne och genre som avgörande faktorer för essäuppgifter. Detta kan tyckas lite märkligt då innehållet ju måste anses fundamentalt för provens validitet. T.ex. torde en indirekt mätning av ordförståelse ge helt andra resultat än en dylik mätning av dispositionsförmåga, på samma sätt som en direkt mätning typen ”Ditt bästa sommarminne” mäter något annat än ”Argumentation för högre bilskatt”.

I en nylig utvärdering av *GRE Analytical Writing Measure* (AW) anges provets innehåll som en primär indikator på mätningens validitet.

The validity of the AW can be demonstrated by: 1) linking the content of the test to necessary skills identified by potential users in the graduate community, 2) establishing the construct that is intended to be measured, and 3) conducting special validity studies. (ETS, 2004, s. 4)

När sedan färdigheterna identifierats förutsätts provdeltagarens utförande av dessa färdigheter ge validitet till mätningens innehåll.

Interviews with graduate faculty have consistently identified critical thinking and writing skills as important for success in graduate school. The two tasks that comprise the Analytical Writing measure are both considered essential in many fields of graduate study. These two tasks are complementary in that the first requires the writer to construct his/her own argument about an issue, and the second requires a critique of someone else's argument by assessing its claims. *The structure of the test can thus be shown to have content validity because the test assesses skills identified by the graduate community as essential for success in many fields of graduate work* [min. kurs.]. (ETS, 2004, s. 4.)

Att provets innehållsbeskrivning ligger i linje med avnämarnas beskrivning av de färdigheter som ger framgång i studier kan visserligen vara en önskvärd del av mätningens innehållsvaliditet. Dock är detta inte ett självklart mått på validiteten och är i sig inte tillräckligt som bevis för denna. Efter att man ringat in den förmåga man avser mäta måste man länka denna till kriterier som kan indikera denna förmåga hos provdeltagaren. Avnämarnas kategorisering av förmågan eller den handling som provdeltagaren förväntas utföra kan då förstås bedömas som goda kriterier.

Vad gäller de indirekta mätningarna är detta problem inte lika aktuellt då dessa fokuserar på mer specifika kunskaper och färdigheter, t.ex. att kunna "identifiera felaktigt ordval" eller "förbättra en mening". De direkta mätningarna, däremot, griper över betydligt mer komplexa förmågor, som t.ex. *the New SAT: the Writing Section*, på vars hemsida man kan ana en liknande "cirkelbevisning" av innehållsvaliditeten.

You'll be asked to develop a point of a view on an issue, using reasoning and evidence, based on your own experiences, readings, or observations, to support your ideas. [---] The short essay measures your ability to: Organize and express ideas clearly; Develop and support the main idea; Use appropriate word choice and sentence structure. (www.collegeboard.com.)

Även om en resonerande uppsats säkerligen mäter delar av en provdeltagares resonerande förmåga och det korta formatet säkerligen fordrar förmågan att disponera, borde kraven på innehållsvaliditeten, även från allmänheten, föranleda en tydligare redovisning av samband.

För Högskoleprovets del är det nödvändigt att ställa det nuvarande provets innehåll i relation till en eventuell förändring av provet, t.ex. införandet ett skrivprov. Bl.a. skulle det vara intressant att analysera innehållet i de verbala delarna i Högskoleprovet (ORD, LÄS och ELF) och jämföra deras korrelationer med olika kriterier för skrivförmåga. Beroende på vilken definition man ger begreppet mäter vi kanske redan nu olika delar av skrivförmåga. Likaså skulle man behöva studera de befintliga delprovets relation till en mängd innehållsmässigt olika skrivuppgifter. Även olika aspekter på mätsäkerhet torde ge upphov till studier av uppgifternas innehåll.

### **7.3 Ett nytt delprov ger nya utmaningar**

Om det skulle visa sig fruktbart att införa ett skrivprov i Högskoleprovet skulle projektet stå inför en mängd nya utmaningar och forskningsområden. En första fråga som aktualiseras är det totala provets innehåll och sammansättning. En uppenbar konsekvens om man helt sonika tillför ett skrivprov till det befintliga provet är att den verbala delen skulle utökas till fyra delprov medan den kvantitativa

består av två, vilket kan ge oönskade följder för mätningen. Det har, vad gäller denna balans, tidigare framförts önskemål att den verbala delen ska beskärmas för provet då förmodas bli mindre känsligt för upprepat provtagande (Högskoleverket, 2000, s. 27).

Om förutsättningarna för ett skrivprov skulle visa sig gynnsamma finns det dock även andra anledningar att se över provsammanställningen. Inte minst kräver de administrativa aspekterna på Högskoleprovet grundliga undersökningar för att fastställa allt från provtiden till utprövningsförfarande. Om skrivprovet helt eller delvis ska bestå av direkta mätningar måste bl.a. frågor om rättning och hur denna ska finansieras att besvaras.<sup>14</sup> En viktig del i utvecklingsarbetet blir då även att konstruera rättningssmallar och definiera svårighetsgrader samt utforma förfaranden så att kvaliteten i rättningen kan säkras.

I samband med de administrativa aspekterna måste man också överväga för- och nackdelar med att datorisera provet. Ett datoriserat prov skulle t.ex. kunna effektivisera och minska kostnaderna för rättning av essäuppgifter, men skulle samtidigt öka kostnaderna för provgenomförandet samt ställa nya krav ur sekretesshänseende. Det blir också relevant att undersöka vilka hjälpmedel som ska erbjudas provdeltagare med funktionshinder. Förutom de hjälpmedel som vi hittills erbjuder, längre provtid, förstord text och punktskrift kan en funktionshindrad provdeltagare behöva t.ex. en skribent för att kunna genomföra en essäuppgift.

Sammanfattningsvis måste beslutet om införandet av ett skrivprov i Högskoleprovet föregås av såväl principiella diskussioner, med utgångspunkt i Högskoleprovets syfte, funktion och ekonomiska villkor, som gedigna fältstudier och utprövningsförfaranden. Detta för att Högskoleprovet ska kunna möta både avnämarnas behov och allmänhetens förväntningar om ett legitimt och mätsäkert prov.

---

<sup>14</sup> Skolverket har i *Mellan myt och verklighet*, 2004, skissartat uppskattat den ekonomiska kostnaden av enbart rättningen till 12 miljoner kronor per år (utifrån antagandet att 100 000 genomför provet per år) (s. 57).

## Litteraturförteckning

- Breland, H., Bridgeman, B. & Fowles, M. (1999a). *Writing Assessment in Admission to Higher Education: Review and Framework*, College Board Report No. 99-3; GRE Board Research Report No. 96-12R. New York: College Entrance Examination Board.
- Breland, H., Kubota, M. & Bonner, M. (1999b). *The Performance Assessment Study in Writing: Analysis of the SATII: Writing Subject Test*, College Board Report No. 99-4. New York: College Entrance Examination Board.
- Breland, H., Kubota, M., Nickerson, K., Trapani, C. & Walker, M. (2004). *New SAT Writing Prompt Study: Analyses of Group Impact and Reliability*. CollegeBoard Research Report No 2004-1; ETS RR-04-03. New York: College Entrance Examination Board.
- Bridgeman, B. (1991). "Essays and multiple-choice tests as predictors of college freshman GPA", *Research in Higher Education*, Vol. 32, No. 3.
- Cooper, P. (1984). *The assessment of Writing Ability: A Review of Research*, GRE Board Research Report GREB No. 82-15R; ETS Research Report 84-12. Princeton, NJ: Educational Testing Service.
- Högskoleverket (2000). *Högskoleprovet. Gårdagens mål och framtida inriktning*, Högskoleverkets rapportserie 2000: 12 R. Stockholm.
- Lawrence I., Rigol, G., Van Essen, T. & Jackson, C. (2003). *A historical Perspective on the Content of the SAT*, College Board Research Report No. 2003-3; ETS RR-03-10. New York: College Entrance Examination board.
- Powers, D., Fowles, M. & Boyles, K. (1996). *Validating a Writing Test for Graduate Admissions*, GRE Board Professional Report No. 93-26bP; ETS Research Report 96-27. Princeton, NJ: Educational Testing Service.

- Powers, D., Fowles, M. & Welsh, C.(1999). *Further Validation of a Writing Assessment for Graduate Admissions*, GRE Board Research Report No 96-13R; ETS Research Report 99-18. Princeton, NJ: Educational Testing Service.
- Powers, D., Fowles, M. & Willard, A. (1993). *Direct Assessment, Direct Validation? An Example From the Assessment of Writing*, ETS Research Report 93-41. Princeton, NJ: Educational Testing Service.
- Reuterberg, S-E. (2003). *Vilken betydelse har utländsk bakgrund för resultatet på högskoleprovet?* Högskoleverkets rapportserie 2003:23 R. Stockholm.
- Skolverket (2004). *Mellan myt och verklighet. Underlag utarbetat med anledning av SOU 2004:29 Tre vägar till den öppna högskolan*. Stockholm.
- Stage, C. & Ögren, G. (2004). *Högskoleprovet våren och hösten 2003. Provdeltagargruppens sammansättning och resultat*, PM Nr 193, 2004. Umeå: Enheten för pedagogiska mätningar.
- The GRE Analytical Writing Measure: An Asset in Admissions Decisions*, (2004). Princeton, NJ: Educational Testing Service.
- The Swedish national aptitude test: a 25-year testing program current status and future development* (2002). Stockholm: Högskoleverket.
- Tre vägar till den öppna högskolan*, SOU 2004:29. Stockholm.
- Vilken betydelse har utländsk bakgrund för resultatet på högskoleprovet? II*, Högskoleverkets rapportserie 2003:23. Stockholm: Högskoleverket.
- Wolming, S. (2000). *Validering av urval*. Umeå universitet, Umeå.

## **Elektroniska källor**

[www.ets.org](http://www.ets.org)

[www.collegeboard.com](http://www.collegeboard.com)

[www.gmac.com](http://www.gmac.com)

[www.gre.org](http://www.gre.org)

[www.kaptest.com](http://www.kaptest.com)

[www.lsat-center.com](http://www.lsat-center.com)