

Gustaf B. Skar & Kjell Lars Berge

Elevens skriveförmåga och texters kvantitativa egenskaper

Trondheim, mai 2017



Skrivesenterets skriftserie 1
Elevers skrivfrmga och texters kvantitative egenskaper
Gustaf B. Skar & Kjell Lars Berge
© 2017 Forfatterne og Skrivesenteret



Skrivesenteret

Nasjonalt senter for skriveopplring
og skriveforskning

www.skrivesenteret.no

ISBN 978-82-93194-18-7 (pdf)

Omslagsfoto: Skrivesenteret
Layout: Skrivesenteret

Forord

Nasjonalt senter for skriveopplæring og skriveforskning skal bidra til at den nasjonale utdanningspolitikken blir iverksatt og gjennomført slik at alle barn, unge og voksne kan få en likeverdig og tilpasset opplæring av høy kvalitet i et inkluderende fellesskap. I oppdragsbrevet som Skrivesenteret mottar fra Utdanningsdirektoratet står det at senteret skal «samle, systematisere og formidle resultater fra forsknings- og utviklingsarbeid til sektorene» og at senteret «skal kjenne til og bruke et bredt og relevant eksisterende kunnskapsgrunnlag i sitt arbeid». Som et ledd i å utføre disse oppdragene gjennomfører Skrivesenteret utviklingsarbeid og lager kunnskapsoversikter. Disse presenteres i Skrivesenterets skriftserie, som består av tekniske rapporter og meldinger.

I denne tekniske rapporten, «Elevens skrivförmåga och texters kvantitativa egenskaper», har Gustaf B. Skar (Skrivesenteret/NTNU) og Kjell Lars Berge (Skrivesenteret/UiO) gjennomført kvantitative analyser av elevtekster for å så korrelere disse mot tekstenes karakterer. Til grunn for analysene ligger 50 elevtekster skrevet av 25 elever som svar på to oppgaver, en «utforskende» og en «forestillende». Hver tekst er vurdert av åtte vurderere. Funnene indikerer høyt samsvar mellom karakter og tekstlengde og lange ord på de sakpregede (utforskende) tekstene og høyt samsvar mellom karakter og ordvariasjon på forestille seg-tekstene. Korrelasjonsanalysene viser også at elevene ikke skriver like langt i de to oppgavene de besvarer, men at det er et høyt samsvar i ordvariasjon i de to oppgavene. I rapporten blir resultatene diskutert.

Skar og Berge har forfattet rapporten og Aasen har vært redaktør. Pia Farstad Eriksen har transkribert elevtekster. Aasen og Skar er redaktører for Skrivesenterets skriftserie.

Arne Johannes Aasen
Senterleder, redaktør

Gustaf B. Skar
Prosjektleder, redaktør

Introduktion

Denna tekniska rapport ägnar uppmärksamhet åt relationen mellan resultat på skriftliga uppgifter och några kvantitativa textmått, nämligen textlängd, ordvariation och syntaktisk variation. Syftet med undersökningen är att identifiera variabler som eventuellt samvarierar med betyg, något som i sin tur kan bidra till att förklara om det finns språkbruksdrag som verkar bidra till att texter uppfattas vara av olik kvalitet, eller att mer eller mindre utbyggt innehåll. Ytterligare ett syfte är att undersöka elevstabilitet, dvs. i vilken utsträckning elevers texter uppvisar liknande drag oberoende av uppgift.

Resultaten är viktiga av flera skäl. För det första kan de ses i ljuset av den pågående diskussionen om reliabilitet i bedömning av texter; vi behöver alltså kunskap om i vilken utsträckning externa variabler kan bidra till att förutsäga mänsklig bedömning av texter. Vi behöver också kunskap om i vilken utsträckning olika ideal ligger till grund för bedömning av texter som är svar på olika typer av uppgifter. En undersökning av samvariationen mellan betyg och kvantitativa textmått kan indikera om bedömare urskiljer olika aspekter som olika viktiga vid olika tillfällen. Slutligen kan resultaten bidra med viss kunskap om uppgifters inverkan på elevers texter.

Texterna som ingår i analysen har skrivits av elever i årskurs 8 (dvs. 8. trinn), som svar på uppgifter i *de nasjonale utvalgsprøvene i skrivning* som grunnleggende ferdighet. Alla elever har skrivit två texter, som i sin tur är svar på två typer av uppgifter: en som ber eleven att "utforska" ett tema och en som ber eleven att "föreställa sig" något. Uppgifterna har tagit sin utgångspunkt i modellen "Skrivehjulet" (Berge, Evensen, & Thygesen, 2016) och representerar två olika sätt att skriva och två olika syften med skrivandet. Det första syftet handlar om att presentera kunskap på ett sätt så att den kan synas i sömmarna, och det andra om att använda skrivande för att skapa föreställningsvärldar (se vidare <http://www.skrivesenteret.no/ressurser/skriveproven/>).

Resultaten på proven är en sammanvägning av bedömning av sex deldimensioner, eller bedömningsområden (se mer nedan). Kvantitativa textmått relaterar ofta till det bedömningsområde som i detta sammanhang kallas språkbruk. I den här rapporten har vi dock utgått från helhetsomdömet. För att förstå hur ett eller flera kvantitativa mått relaterar till bedömningen av elevens samlade skrivförmåga måste vi av uppenbara skäl ta hänsyn till bedömningen av alla områden och inte begränsa oss till det som tydligast är relaterat till de textaspekter som de kvantitativa textmått relaterar till. Vidare kan man tänka sig att textlängd per se inte är relaterat till viss nivå av stilvariation, utan fångar i vilken utsträckning innehållet är utbyggt.

Material

Underlaget för studien kommer från elevtexter som samlats in och bedömts inom ramen för projektet *nasjonale utvalgsprøver i skrivning* (Skar & Iversen, 2016). Mer specifikt rör det sig om 50 elevtexter skrivna av 25 elever från årskurs 8. Texterna, som omfattar 19 386 ord, har skrivits som svar på följande två uppgifter:

Uppgift 1584: Skriv en tekst der du utforsker årsaker til at røyking var mer akseptert i samfunnet før enn nå. Du skriver teksten til et hefte klassen din lager om rusmidler.

Uppgift 1585: Forestill deg at du er på vei hjem en sen høstkveld og ser et merkelig lys. Skriv en tekst der du forteller om hva du opplevde denne kvelden. Tenk deg at du skal lese teksten opp for klassen, kanskje mens dere sitter i et mørkt rom med stearinlys.

Samtlige texter har bedömts av åtta specialtränade bedömare i samband med ett forskningsprojekt (Skar & Jølle, under utg.). (Texterna har efter genomförd bedömning normaliserats vad gäller stavning, men i övrigt inte bearbetats.) De åtta bedömarna har varit en del av den så kallade bedömarpanelen, som bestått av lärare från hela Norge och som under flera år bedömt elevtexter insamlade i samband med Nasjonale utvalgsprøver i skrivning. Att varje text är läst av åtta bedömare ökar resultatens säkerhet.

Varje text har bedömts på sex bedömningsområden/-skalor: *kommunikation*, *innehåll*, *textstruktur*, *språkbruk*, *stavning* och *interpunktion* (läs mer i Skar & Iversen, 2016). Varje skala har varit femgradig och det mittersta skalsteget (3) har haft innebörden "som förväntat för årskursen". Med detta menas att kvaliteten på texten motsvarar det som lärare menar är rimligt att förvänta av elever efter fullgod skrivundervisning (Evensen, Berge, Thygesen, Matre, & Solheim, 2016).

Totalt har varje text erhållit 48 bedömningar (6 områden * 8 bedömare). Inalles omfattar materialet 2 400 bedömningar. I den här rapporten används det samlade resultatet för varje elevtext (dvs. genomsnittet av de 48 bedömningarna). Detta resultat, som vi kallar "betyg", är framtaget genom en statistisk metod som kontrollerar för kända aspekters inverkan på resultaten (resultatet kallas med en teknisk term "fair score", är frukten av så kallad many-facet Rasch-analys och kontrollerar t.ex. för bedömarstränghet; Skar & Iversen, 2016). Korrelationen mellan uppgiftsresultaten var stark med $r = 0,85$ och 13 elever fick samma betyg på båda uppgifter.

Dataanalys

Texterna har analyserats i webapplikationen «Lix.se», som genererar nio mått: antal meningar, antal ord, antal

långord, genomsnittlig meningslängd, andel långord, läsbarhetsindex, type/token-ratio, ordvariationsindex och ordvariationsratio. I denna undersökning har vi använt följande mått:

- Antal ord (o), som används för att undersöka relationen mellan textlängd och betyg.
- Genomsnittlig meningslängd (LM), som används för att undersöka relationen mellan meningslängd och betyg. (Mättet antal meningar är uteslutet ur analysen eftersom det inte bidrar med ytterligare information.)
- Antal långord (L ; andel långa ord: $Lo = L/o * 100$), som används för att undersöka relationen mellan potentiellt ovanliga och specialiserade ord och betyg. (Mättet *andel långord* är uteslutet ur analysen eftersom det inte med nödvändighet bidrar med ytterligare information.)
- Läsbarhetsindex ($LIX = LM + Lo$), som används för att undersöka relationen mellan den antagna läsbarheten och betyg.
- Ordvariationsindex ($OVIX, [\log(\text{tokens}) / \log(2 - (\log(\text{types}) / \log(\text{tokens})))]$), som kontrollerar för textlängd, som används för att undersöka relationen mellan ordvariation och betyg. (Måtten type/token-ratio och ordvariationsratio är uteslutna ur analysen eftersom de inte bidrar med ytterligare information.)

Stor ordvariation och många långord indikerar ett specialiserat register (jfr Hultman & Westman, 1977, s. 56, 78; Östlund-Stjärnegårdh, 2002, s. 87ff), medan långa texter kan indikera ett utbyggt innehåll. En vanlig tolkning av läsbarhetsindex är följande:

- < 30: Mycket lättläst, barnböcker
- 30–40: Lättläst, skönlitteratur, populärtidningar
- 40–50: Medelsvår, normal tidningstext
- 50–60: Svår, normalt värde för officiella texter
- > 60: Mycket svår, byråkratspråk

Det skall emellertid sägas att det inte råder något kausalsamband mellan någon kvantitativ textvariabel och kvalitativa textegenskaper. Ett av skälen är att textvariablerna bygger på analys av ytliga textdrag som inte tar hänsyn till textens innehåll och den mening som läsaren är tänkt att ta ur texten. En text kan mycket väl ha egenskaper som vid genomläsning skulle få den att framstå som av lägre kvalitet än vad måtten indikerar. Detta gäller t.ex. textlängd som ju kan vara resultatet av nonsenspartier, eller genomsnittlig meningslängd som kan öka väsentligt om författaren inte sätter ut punkt. Med detta sagt är inte heller syftet med denna rapport att fastslå hur kvalitativa egenskaper samvarierar med betyg. Det blir en fråga för andra typer av undersökningar.

För att svara på frågan om vilka textegenskaper som tycks samvariera med betyg och hur elevstabiliteten ser ut, har vi genomfört enkla bivariata korrelationsanalyser. I analysen har de stilistiska måtten korrelerats med betygen. Vidare har de stilistiska måtten för texter för respektive uppgiftstyp korrelerats med varandra.

Resultat

I tabell redovisas den deskriptiva statistiken för betyg och textmått. De 25 eleverna fick i genomsnitt 2,68 poäng på U1584 ("utforska") och 2,90 poäng på U1585 ("föreställa sig"). Resultaten var alltså något högre på den senare uppgiften och nära gränsen för "förväntad nivå". Detta kan tolkas som att eleverna i allmänhet bättre behärskade skrivhandlingen att föreställa sig än skrivhandlingen att utforska. Vi ser också att det var viss spridning bland eleverna med standardavvikelser på 0,73 och 0,80 för utforska- respektive föreställa sig-uppgiften. Detta, tillsammans med informationen om minimi- och maximumvärden, anger att skillnaderna mellan enskilda elever var tämligen stor på båda uppgifter. På utforska-uppgiften, för att ta den som exempel, fick eleven med lägst betyg 1,2 poäng (dvs. långt under "förväntad nivå") medan eleven med högst betyg fick 3,94 poäng.

Textmåtten visar att det fanns skillnader relaterade de båda uppgifterna. Till exempel var genomsnittliga textlängden cirka 280 ord för de utforskande U1584-texterna, medan den var cirka 494 ord för de föreställande U1585-texterna. I genomsnitt producerade eleverna därmed omkring 76 % mer text på föreställa sig-uppgiften. Ordvariationen var något lägre i de utforskande texterna (42,57 mot 45,94 i föreställa sig-texterna), vilket indikerar att eleverna haft ett rikare språk i föreställa sig-texterna. LIX-värdet var däremot högre i utforska-texterna (28,16 mot 21,32), något som antyder att eleverna inte bara skrivit längre meningar utan använt relativt sätt längre ord och därmed – kanske – mer specialiserade begrepp (vilket ju skulle vara rimligt med tanke på att U1584 bad eleverna att skriva en sakttext). I tabell 1 framgår vidare av standardavvikelser samt maximum- och minimimått att den individuella variationen mellan texterna var omfattande.

Vänder vi oss till tabell 2, kan vi notera starka och signifikanta samband mellan betygen på de utforskande texterna (U1584) och antal ord ($r = ,72$) och antal långord ($r = ,75$). Sambandet mellan betyg och ordvariation, däremot, är svagt och icke-signifikant ($r = ,36$). Korrelationens riktning indikerar emellertid att ökad ordvariation hänger ihop med högre resultat. Det motsatta förhållandet tycks gälla för resultat och genomsnittlig meningslängd; korrelationen mellan betyg och meningslängd är negativ. Även här är dock korrelationen svag och icke-signifikant ($r = -,27$).

Tabell 1. Deskriptiv statistik: provresultat och stilistikanalyser

		N	Medel	S.E.	S.D.	Min	Max
Betyg	U1584	25	2,68	0,15	0,73	1,20	3,94
	U1585	25	2,90	0,16	0,80	1,54	4,32
U1584	Ord	25	280,64	25,98	129,92	64,00	574,00
	Men.längd	25	16,51	1,19	5,96	7,85	32,00
	Långord	25	33,04	3,61	18,07	3,00	89,00
	LIX	25	28,16	1,12	5,60	18,00	42,00
	OVIX	25	42,57	1,08	5,38	28,98	49,93
U1585	Ord	25	494,80	48,38	241,91	45,00	1110,00
	Men.längd	25	12,25	0,86	4,31	6,19	22,64
	Långord	25	45,96	4,31	21,55	12,00	107,00
	LIX	25	21,32	0,99	4,97	14,00	31,00
	OVIX	25	45,94	1,12	5,58	35,29	55,72

Notera. U1584: utforskande, U1585: föreställande. N: antal texter, ord: antal ord, men.längd: genomsnittlig meningslängd, långord: antal långord, lix: läsbarhetsindex, ovix: ordvariationsindex.

För föreställa sig-texterna (U1585) visar analyserna ett svagt och icke-signifikant samband mellan betyg och antal ord ($r = ,22$) och antal långord ($r = ,36$). Däremot är sambanden mellan betyg och ovix såväl någorlunda starkt som signifikant med $r = ,53$. Tendensen för meningslängd är den samma som för U1584. Korrelationen mellan betyg och LIX är i det närmaste obefintlig (vilket också gäller för U1584).

I tabell 3 redovisas korrelationsanalyser av de textinter-
na variablerna. Som resultaten indikerar är sambandet mellan den viktigaste indikatorn för de utforskande texterna (U1584) (antal ord) och den viktigaste indikatorn för de föreställande texterna (U1585) (OVIX) svagt. Det innebär att textmängden på utforska-uppgiften inte tydligt hänger samman med ordvariationen på föreställa sig-uppgiften. I sin tur innebär det att elever som skriver långt på utforska-uppgiften inte nödvändigtvis skriver varierat på föreställa sig-uppgiften. Om vi läser tabellen diagonalt, dvs. noterar korrelationerna mellan ord-ord, genomsnittlig meningslängd-genomsnittlig meningslängd osv., ser vi att alla korrelationer både är signifikanta och antingen moderata eller starka. Ser vi på de tydligaste stilistkindikatorerna – LIX och OVIX – kan vi notera att elever som producerar texter med högt LIX och OVIX på den ena uppgiften gör det samma på den andra.

Tolkning

Sambandet mellan betyg och antalet ord och antalet långord är starkt för utforska-uppgiften (U1584). Detta indikerar att bedömarna lagt stor vikt vid innehållets omfattning och precision i de utforskande texterna. Ser vi på minimivärdet för texter på U1584, kan det noteras att antalet ord (64) motsvarar cirka fem rader text med standard typografi (Times New Roman, 12 punkter). Det är lätt att tänka sig att en text som är så kort svårligen kan utforska orsaker till att rökning var mer accepterat förr än nu. Med stor sannolikhet innehåller den inte heller många, långa precisa begrepp. I förlängningen pekar resultaten i riktning mot att varken ytliga kunskaper eller vardagligt språk kan användas för att lösa uppgiften på ett för bedömarna tillfredsställande sätt.

Sambandet mellan betyg och antal ord och långord på U1585 ("föreställa sig") är svagt och icke-signifikant. Det indikerar att texternas omfattning eller användning av specialiserade ord i sig inte skiljer svaga och goda elevtexter. Studerar vi medelvärdena och spridningen för textlängd på de bägge uppgifterna verkar resultatet rimligt. Eleverna skriver undantagslöst betydligt längre texter på U1585 och har, sannolikt, få svårigheter att hålla innehållet relevant. Den stilistiska indikator som

Tabell 2. Korrelationsanalyser

	Ord	Men.längd	Långord	LIX	OVIX
Betyg på U1584	,72**	-,27	,75**	-,07	,36
Betyg på U1585	,22	-,26	,36	-,07	,53**

Notera. Ord: antal ord, men.längd: genomsnittlig meningslängd, långord: antal långord, LIX: läsbarhetsindex, OVIX: ordvariationsindex. ** $p < ,01$

Tabell 3. Korrelationer, textvariabler

		U1584				
		Ord	Men.längd	Långord	LIX	OVIX
U1585	Ord	,50*	-,13	,33	-,32	,10
	Men.längd	-0,29	,80**	-0,20	,81**	-0,13
	Långord	,55**	-0,01	,45*	-0,08	0,29
	LIX	-0,22	,70**	-0,06	,80**	0,01
	OVIX	0,32	-0,22	,45*	0,09	,71**

Notera. Ord: antal ord, men.längd: genomsnittlig meningslängd, långord: antal långord, lix: läsbarhetsindex, ovix: ordvariationsindex. * $p < ,05$; ** $p < ,01$

däremot har samband med betygen är språklig variation (ovix). Även om sambandet inte är alltför starkt, indikerar det att elever som skriver texter med varierat språk i föreställa sig-texter i högre grad bedöms motsvara uppgiftens krav på att skapa underhållande berättelser. Det är också rimligt att anta att eleverna i föreställa sig-texterna i högre utsträckning kan använda ett vardagligt språk. Att utforskande saktexter kräver andra språkliga resurser än föreställande vardagstexter är känt sedan tidigare (se t.ex. Berman & Nir-sagiv, 2007; Kamberelis, 1999; Kamberelis & Bovino, 1999).

Undersökningen av korrelationerna mellan textvariablerna indikerar att elevers prestationer inte enkelt låter sig förutsägas med de undersökta textmått (korrelationen mellan uppgiftsresultaten antyder att resultatet som sådant är en långt bättre prediktor). Detta är ett återkommande fynd i tidigare forskning (t.ex. Berge & Skar, 2015; Bouwer, Béguin, Sanders, & van den Bergh, 2015). Vad denna undersökning visar, emellertid, är att eleverna intressant nog tycks vara någorlunda konsistenta och kanske i mindre utsträckning än önskvärt lyckas anpassa språkbruket till den givna uppgiften. Ett sådant resonemang skulle förklara att det är lägre korrelation mellan antal ord på de båda uppgifterna än för ovix; elever kan producera olika mycket innehåll om olika ämnen – en innehållsproduktion som dessutom

tycks viktas olikt – men har tillgång till ungefär samma stilregister oberoende av uppgift. En sådan tolkning skulle innebära följande två tentativa och inte särskilt kontroversiella slutsatser: eleverna i det här materialet kan olika mycket om olika ämnen och har tillgång till ett och samma register oberoende av uppgift. Den sista slutsatsen, skulle innebära att det inte bara är uppgiften i sig som avgör om eleven kan variera sitt språk eller inte.

Avslutning

Resultaten ovan har viktiga implikationer. För det första - och viktigast - implicerar resultaten att vilka variabler som samvarierar med betyg är beroende av uppgift (realiserat som skrivhandling + tema). *Bedömnarna verkar ha använt olika strategier för att bedöma texterna.* De tycks ha tillskrivit innehållslig relevans och precision hög betydelse på uppgifter som ber om utforskande skrivhandlingar, och estetiska dimensioner hög betydelse på uppgifter som ber om föreställande skrivhandlingar.

För det andra indikerar resultaten att elever utan svårighet kan skriva tämligen omfattande på "föreställa sig-uppgifter".

För det tredje, och kanske viktigast, visar resultaten att inget av de kvantitativa språkbruksmåten som använts i denna undersökning kan användas för att tillfredsställande göra förutsägelser om elevers samlade skrivförmåga (så som mätt genom två uppgifter). Fortsatta studier baserade på större material är nödvändiga.

Litteratur

- Berge, K. L., Evensen, L. S., & Thygesen, R. (2016). The Wheel of Writing: a model of the writing domain for the teaching and assessing of writing as a key competency. *The Curriculum Journal*, 27(2), 172–189. <http://doi.org/10.1080/09585176.2015.1129980>
- Berge, K. L., & Skar, G. B. (2015). *Ble elevene bedre skrivere? Intervensjonseffekter på elevers skriveferdigheter og skriveutvikling* (Rapport 2). Trondheim: Høgskolen i Sør-Trøndelag, Avdeling for lærer- og tolkutdanning.
- Berman, R. A., & Nir-sagiv, B. (2007). Comparing Narrative and Expository Text Construction Across Adolescence: A Developmental Paradox. *Discourse Processes*, 43(2), 79–120. <http://doi.org/10.1080/01638530709336894>
- Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, 32(1), 83–100. <http://doi.org/10.1177/0265532214542994>
- Evensen, L. S., Berge, K. L., Thygesen, R., Matre, S., & Solheim, R. (2016). Standards as a tool for teaching and assessing cross-curricular writing. *The Curriculum Journal*, 27(2), 229–245. <http://doi.org/10.1080/09585176.2015.1134338>
- Hultman, T. G., & Westman, M. (1977). *Gymnasist-svenska*. Lund: Liber Läromedel.
- Kamberelis, G. (1999). Genre development and learning: Children writing stories, science reports, and poems. *Research in the Teaching of English*, 33(4), 403–460. Retrieved from <http://www.jstor.org/stable/40171463>
- Kamberelis, G., & Bovino, T. D. (1999). Cultural artifacts as scaffolds for genre development. *Reading Research Quarterly*, 34(2), 138–170. Retrieved from <http://www.jstor.org/stable/748162>
- Skar, G. B., & Iversen, J. M. (2016). *Nasjonale utvalgsprøver i skriving 2015. Teknisk rapport*. Trondheim: Nasjonalt senter for skriveopplæring og skriveforskning.
- Skar, G. B., & Jølle, L. (under utg.). Teachers as raters: investigation of a long term writing assessment program.
- Östlund-Stjärnegårdh, E. (2002). *Godkänd i svenska?: Bedömning och analys av gymnasieelevers texter*. (Doktorsavhandling, Uppsala universitet). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-1967>