

En gyldig vurdering av elevers skrivekompetanse?

Lars Sigfred Evensen

Vurdering er en sentral kunnskapsutfordring for skriveledere og skriveforskere. Utfordringen er kompleks og omfatter både faglige, politiske, etiske og organisatoriske spørsmål (Evensen, 2009). Fra et faglig ståsted må vi for eksempel vurdere på måter som sikrer gyldighet. I politisk sammenheng er vurdering et spørsmål som påvirkes av ikke bare politiske føringer i form av læreplaner, men også av offentlighetens drakamper mellom ulike interesser, ideologier og standpunkter. I tillegg påvirkes læreres vurdering av organisatoriske hensyn – til utdanningssystemets oppbygging, til skolers virkemåte og til læreres arbeidssituasjon.

I denne artikkelen skal jeg se nærmere på kun ett av de rent faglige spørsmålene: Hvordan kan vi sikre at vår vurdering blir gyldig, med hensyn til både å vurdere det vi tror vi vurderer, og samtidig gjøre dette på en måte som ikke innsnevrer det fagområdet vi vurderer i urimelig grad? I skrivende stund er det bestemt av norske myndigheter at et representativt utvalg elever på 5. og 8. årstrinn i grunnskolen skal ha årlige skriveprøver, men det er ikke endelig bestemt hvordan disse prøvene skal utformes. Jeg skal delta i utviklingen av prøvene og bruker derfor de gyldighetsproblemerne det innebærer som mitt illustrerende eksempel.¹

1 Den innsikten som blir presentert i denne artikkelen, er i stor grad utviklet i samarbeid med langvarige prosjektkolleger som Kjell Lars Berge, Rolf Fasting, Ragnar Thygesen og Wenche Vagle. Denne gruppa utviklet de nasjonale skriveprøvene i 2005. Når jeg i artikkelen referer til en ekspertgruppe, er det denne gruppa jeg sikter til.

I Smidt, J., Folkvord, I. og Aasen, A.J. (red.) (2010).

Rammer for skriving. Om skriveutvikling i skole og yrkesliv. Trondheim: Tapir Akademisk Forlag.

Hva forstår vi med «skrivning»?

For å kunne vurdere skrivekompetanse på en gyldig måte er den første forutsetningen at vi presiserer hva vi forstår med skrivning. Kjernebegrepet «skrivning» må altså defineres. Det er mange måter å gjøre det på, og den vanligste er å regne opp skriftas ressurser av tegn, dens semiotiske bestanddeler – som håndskrift, ortografi, tegnsetting, ordforråd, grammatikk og tekststruktur. Skrivekompetanse kan da defineres som en elevs dyktighet til å bruke disse bestanddelene som uttrykksverktøy på en kulturelt akseptabel måte. Sagt på en annen måte: Hvilke semiotiske verktøy har en elev i den verktøykassen som hun er i stand til å bruke med en viss kyndighet? Den internasjonalt sammenlignende IEA-undersøkelsen av skriveferdighet (se Gorman, Purves & Degenhart, 1988) valgte en bred variant av denne strategien. I lys av de siste tiårenes skriveforskning kan vi imidlertid se at strategien har en vesentlig svakhet – den sier ikke noe om hva skrivning skal tjene til; hva elever skal med sin skrivekompetanse. Vurdert ut fra et sosialt perspektiv på skrivning, mangler altså selve den grunnleggende komponenten i begrepsforståelsen. Derfor er det behov for å utvikle en alternativ strategi.

Da nasjonale skriveprøver ble utviklet i Norge i perioden 2003–2005, valgte ekspertgruppa bak disse prøvene (se Thygesen, Evensen, Berge, Fasting, Vagle & Haanæs, 2007) en alternativ strategi hvor det var nettopp hensikten med skrivning som ble lagt til grunn. Gruppa tok sitt første utgangspunkt i det vi vet om skrivingas historiske opphav og prinsipielle kjerneegenskaper. Flere studier peker på at skriftlig mediering ble brukt svært tidlig til ulike former for korrespondanse. Et vestlig eksempel er Bazermans (2002) studie av det uformelle brevet som opphav til selv noen av de mest formelle moderne sjangrene, som institusjonelle attester og nasjonale charter. Et afrikansk eksempel er Scribner og Coles (1981) studie av Vai-folket, hvor brev benyttes til å holde kontakt mellom fjerntliggende stammer. Det er også flere studier som peker på at skriftlig mediering er benyttet svært tidlig til å dokumentere historiske hendelser, varebeholdninger eller økonomiske transaksjoner. Et velkjent nordamerikansk eksempel i så måte er sletteindianernes bildeskift (Goody & Watt, 1963).

Disse studiene synes å vise at kommunikasjon og dokumentasjon er grunnleggende formål som skrivning har ivaretatt over en flertusenårig lang historisk periode. Og disse hensiktene passer som hånd i hanske

til den skriftlige medieringens kjerneegenskaper. Mens tale i et historisk perspektiv er flyktig, er skrift bestandig. Mens muntlig kommunikasjon har utviklet seg i en situasjon hvor samtalepartnere befinner seg på samme sted samtidig, opphever skriftlig kommunikasjon denne forutsetningen. Når vi skriver, kan vi kommunisere med andre på tvers av lange avstander, og en tekst blir vanligvis skrevet og lest på forskjellig tidspunkt.

Grunnhensiktene gir imidlertid ikke en uttømmende liste over skrivings grunnleggende funksjoner. Det fins studier som synes å vise at det over tid oppstår også avledede hensikter i møtet mellom kommunikasjon og dokumentasjon. For eksempel er det tankevekkende å lese gamle runer som uttrykker bare navnet på den som risset en bestemt rune («NN reist runar þessar»). Vi kan derfor anta at skrivings egenskaper allerede på vikingenes tid inviterte menneskene til å bruke skrivning til det vi med et moderne uttrykk kan kalle identitetsarbeid. Skrivning av runer var en langsom kommunikasjonsform, og sammen med skriftas bestandighet inviterer en slik kommunikasjonsform til refleksjon mens selve skriveprosessen pågår (Evensen, 1997). En slik refleksjon er selvsagt knyttet til også muntlig kommunikasjon, ettersom vi alltid er tilhørere til både andres og egne ytringer, men langsom skrivning og bestandige ytringer inviterer i større grad til ettertenksomhet og metaperspektiv (Olson, 1994). Det er en mekanisme som er viktig i et individuelt lærings- og utviklingsperspektiv, men den har sannsynligvis spilt en rolle også under utvikling av kulturers kollektive identitet. Steinhogde hieroglyfer som er funnet ikke bare fra Farao-dyrkingens Egypt, men også fra de amerikanske Maya- og Aztek-kulturene, hadde gjerne et religiøst og/eller politisk-historisk innhold, og vi må anta at de spilte en viss rolle i utviklingen av kulturell identitet også på et kollektivt plan, selv om skrivning den gang var forbeholdt ei meget lita gruppe innen samfunnets maktapparat.

Skriftas bestandighet gir opphav til også avledede funksjoner som logisk klassifisering og intellektuell etterprøvbarehet. Logisk klassifisering kan ha oppstått først (jf. Goody & Watt, 1963) ettersom dokumentasjon av eiendeler naturlig ble klassifisert etter hvilken type av eiendeler det var snakk om (husdyr, avlinger, utstyr o.l.). Tilsvarende var transaksjoner gjerne av ulike typer (oppgjør, gjeld, utestående fordringer o.l.), noe som gjorde at de ble nedskrevet i ulike deler av et dokument eller i separate dokumenter. Videre kunne forklarende kommentarer til objekter

eller tegninger koble en klassifisering til kommunikasjon av disse. I et moderne kunnskapssamfunn er innholds- eller kunnskapsorganisering blitt en enda mer sentral hensikt med skriving. Alle fag har sine klassifikasjonssystemer og stiller store krav til dokumentasjon. Skriftlig kunnskapsorganisering framstår derfor som et nesten selvsagt aspekt i omtrent enhver faglig virksomhet.

Også i muntlige kulturer blir sentrale historiske, kulturelle, politiske eller personlige fenomener dokumentert – gjennom fortellinger, sagn, myter, sanger, danser eller kollektive ritualer. Dokumentasjonen blir gjerne framført av sjamaner, skalder eller familiens eldste; personer som ofte har lært de viktige «dokumentene» mer eller mindre ordrett. Som kunnskaps- og kulturoverføring mellom mennesker eller generasjoner er en slik praksis svært effektiv. Svakheten ved en muntlig form for overføring er imidlertid at den i liten grad er etterprøvable, eller inviterer til andre former for kritikk. Ettersom kulturbærerne ofte er underlagt den lokale politiske makten, er det videre lett for makthavere å endre elementer i «dokumentene» slik at de passer med aktuelle interesser, agendaer eller oppfatninger. Med skriftas bestandighet endrer dette seg straks skrift blir mer allment tilgjengelig, slik den ble i antikken. Med skriftlige dokumenter blir det mulig for alle lesekyndige som har tilgang til dokumentene å gå tilbake til dem og kontrollere om en framstilling er etterrettelig (Goody & Watt, 1963, s. 322). Med skriftlige dokumenter blir derfor tilgangen til dem også en viktig maktfaktor. Det er denne sammenhengen mellom tilgang til dokumenter og makt som semiotikeren Umberto Eco illustrerer i romanen *Rosens navn*. I denne romanen blir en serie mord begått i et kloster for å hindre tilgang til et dokument som kan true den rådende religiøse doktrinen.

Refleksjon gjennom skriving har imidlertid også en annen mulighetshorisont enn den etterprøvende. Når vi reflekterer eller kritiserer, kan vi begynne å tenke i alternative baner. Dermed kan vi av og til se sammenhenger som vi ikke tidligere har tenkt på:

The magic of writing arises not so much from the fact that writing serves as a new mnemonic device, an aid to memory, as from the fact that writing may serve an important epistemological function. Writing not only helps us remember what was thought and said but also invites us to see what was thought and said in a new way. (Olson, 1994, s. xv)

La meg bruke min egen framstilling som eksempel på dette. Da ekspertgruppa bak de nasjonale prøvene i 2005 valgte en sosialt orientert strategi for å definere kjernebegrepet «skrivning» (Thygesen mfl., 2007), hadde vi ikke noe bevisst perspektiv på forholdet mellom kulturhistoriske grunnfunksjoner og avledede funksjoner. Men flere skriftlige framstillinger av vår begrepsforståelse har invitert til refleksjonsprosesser hvor et slikt perspektiv gradvis har vokst fram. Det er altså ikke alltid slik at vi i skrift bare nedtegner det vi har tenkt ut på forhånd. Snarere kan det være slik at mens vi skriver, tenker vi videre. Og i lykkelige øyeblikk fører denne tenkningen til nye tanker, og kanskje endog til ny innsikt. Gjennom å skrive kan vi faktisk lære.

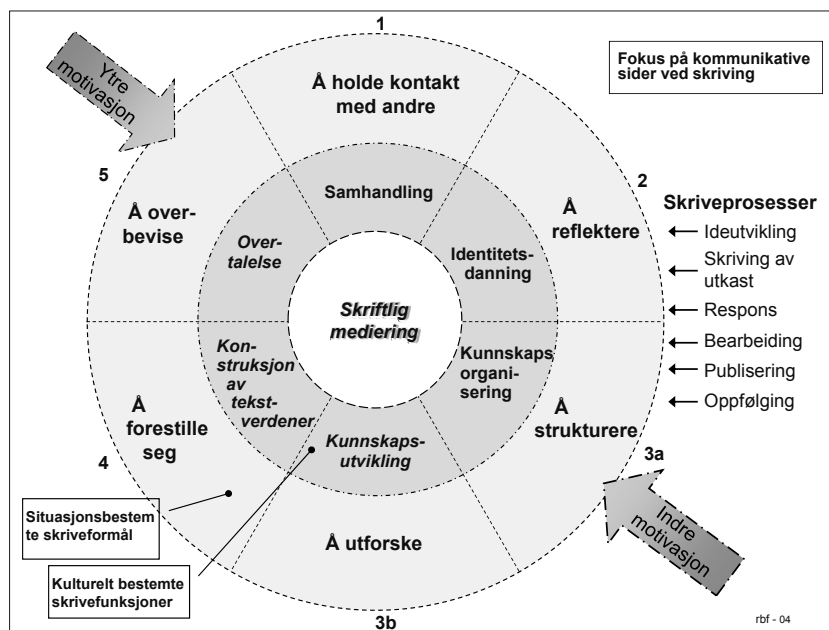
Ikke bare kan vi lære, vi kan også forestille oss noe som aldri har eksistert. «Hva om?», kan vi spørre, eller «Enn hvis?». Slike forestillinger gir et felles opphav til så ulike fenomener som forskeres teorier og dikteres alternative verdener. Riktig nok har slett ikke skriftlig kommunikasjon noe monopol på å fremme kreativ fantasi, men igjen er skriftas bestandighet en støttende egenskap. For å følge norsk språkbruk «filer» vi på våre tekster, og dermed knytter vi oss til håndverkeres tradisjon, hvor kunsten består i å undertvinge seg en motvillig materie som bare høyst motstrebende lar seg temme. Underlig nok er denne filingen et vesentlig aspekt ved all kreativ bruk av skrivning, enten den gjelder estetiske eller teoretiske formål. På dette punktet møtes altså diktning og vitenskap.

Alternative verdener kan også ha betydning for oppfatninger, synspunkter eller verdier. Antikkens retorikk utnyttet denne innsikten. Vi kan forsøke å overbevise andre mennesker om at en annen oppfatning enn den rådende kan være mer gyldig, mer dekkende, mer sannferdig eller mer fruktbar. Vi kan altså argumentere for våre politiske standpunkter eller faglige synspunkter. Antikkens talekunst ble utviklet som et middel for å overtale talernes tilhørere i sosialt betydningsfulle sammenhenger. Dette skjedde i en overveiende muntlig kultur. Skriftlig kommunikasjon tilførte senere noen fortrinn, for gjennom skriftas bestandighet blir som nevnt argumentasjonen etterprøvbare. Antikkens retorer kunne riktignok gi inntrykk av å ha en uovertruffen hukommelse (de hadde selvsagt før den muntlige framføringen lært sin nedskrevne disposisjon utenat), men ikke desto mindre skulle den skrevne argumentasjonen bli forbundet med framveksten av et moderne demokrati. Det er fremdeles først og fremst gjennom en skriftlig argumentasjon at en «taler» kan oppnå «autoritet» (avledet av «autor» = 'forfatter'). Dermed ble alfabe-

tisering av fattige undertrykte også viktig for en moderne frigjøringspedagog som Paulo Freire, som sier at skrivekyndighet *myndig* gjør den som besitter den. Med skrivekyndighet følger enkelt sagt muligheten til å ha innflytelse også utenfor den nærmeste omgangskrets.

Noe som kanskje er overraskende ved min framstilling så langt, er at vi nå gjennom mange ledd faktisk har koblet en avledet hensikt tilbake til den kommunikative grunnhensikten med skriftlig kommunikasjon. Med argumenterende skriving blir konkrete lesere igjen avgjørende deltakere i den skriftlige kommunikasjonen; de blir dermed også avgjørende størrelser i mitt forsøk på å beskrive en vurderingsmessig *gyldig* oppfatning av hva «skrivning» (og «skrift») dreier seg om. Når vi på denne bakgrunnen igjen vender tilbake til den kommunikative grunnhensikten, kan vi begynne å ane konturene av en sirkel som på en eller annen måte er sluttet.

I figuren nedenfor blir de kulturelle skriveformålene framstilt som en sirkel eller et hjul², og leserne vil se at min framstilling så langt har fulgt



Figur 1

2 Den tekniske utforminga av figur 1 og 2 er det Rolf Fasting som står bak.

denne sirkelen hele veien rundt; fra toppen av sirkelen har jeg fulgt solretningen rundt sirkelen, tilbake mot toppen. Et slikt skriftkulturelt hjul nedtoner på ingen måte betydningen av de semiotiske uttrykksredskapene som skrift tilbyr. Tvert imot ser vi at figuren uttrykker den skriftlige medieringen som selve navet i hjulet. Vi ser også at de kulturelt definerte skrivefunksjonene er omgitt av en ytre sirkel. Den skal vi straks drøfte, og derved knytte skriftkulturell historie til levende elever som skriver i en utdanningskontekst.

Fra kulturhistorie til klasserom: Om skriftlige elevytringer

Ulike kulturelt bestemte skrivefunksjoner har i løpet av en lang historie gitt opphav til mange sjangrer. Min framstilling inviterer derfor til et perspektiv hvor vi tillegger sjangrer stor betydning i vurdering av elevers skrivekompetanse. Det er imidlertid ikke noen logisk naturlig følge.

Sjangrer er ikke bare knyttet til bestemte formål som er stadig tilbakevendende (Miller, 1984); de oppstår også i tilknytning til bestemte sosiale grupper som utvikler bestemte måter å uttrykke seg på. Swales (1990) bruker betegnelsen «diskurssamfunn» om slike grupper. Bestemte sjangrer lever altså bare i bestemte sosiale økologier, og klasserommet kan bare i begrenset grad være del av disse ulike økologiene. La meg belyse det opplagte ved dette punktet gjennom litt ytterliggående eksempler. Elever i et klasserom kan ikke med hell lære seg de sjangrene som helsepersonell benytter i sin internkommunikasjon (som f.eks. «epikrise»). De kan heller ikke ha mye håp om å bli dyktige til å skrive lovproposisjoner. Klasserom tilhører selvsagt en økologi hvor det eksisterer egne spesifikke sjangrer, men det forhindrer ikke at en sjangerfokuseret skriveundervisning eller vurdering lett kan bli ugyldig. I tillegg vil det lett kunne binde elever til en liksomverden som mangler genuin motivasjon og dermed med prinsipiell nødvendighet gjør dem til tapere. Det fins altså en viktig teoretisk grunn til at gyldig vurdering av elevers skrivekompetanse ikke bør tuftes på et sjangerperspektiv.

Fins det så et alternativ som er i samsvar med framstillingen ovenfor? Noe av det aller viktigste ekspertgruppa bak de nasjonale prøvene i 2005 gjorde, var å utvikle et slikt alternativ (Thygesen mfl., 2007). Selv om elever og lærere bare har tilgang til et begrenset antall sjangrer fra klasse-

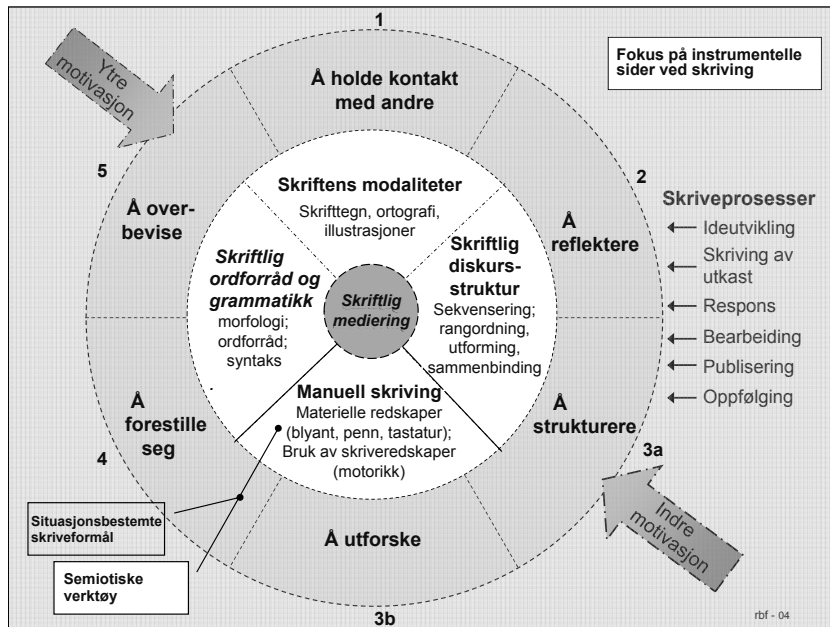
rommet, har de tilgang til alle de samme formålene eller grunnfunksjonene som skriftkulturen ivaretar. Elever kan ytre seg skriftlig for å holde kontakt med andre. De kan bruke skriving til å utvikle sin identitet og autoritet; de kan bruke skriving i sine læreprosesser til både å organisere kunnskap og å utvikle ny kunnskap gjennom refleksjon. Videre kan de skape tekstverdener om noe som *kunne ha vært* eller *burde ha vært*, og de kan skrive for å overbevise andre om at deres synspunkt er verd å deles av andre. Slike ytringer vil etter hvert nærme seg sjangerer, men vil samtidig ikke være låst til en sjangeroppskrift som elever sjelden vil kunne lykkes med. Klasserommet tilbyr rike muligheter for skriftlige ytringer av mange slag, og det er nettopp slike ytringer som vil kunne vurderes på en gyldig måte: Hvor godt ivaretar elevytringen sin hensikt, gitt det alders- og utviklingstrinnet som en elev befinner seg på? Det siste skal jeg vende tilbake til etter neste avsnitt.

Det er dette resonnementet som ligger til grunn for den ytre sirkelen i figur 1. Det er også den ytre sirkelen som forbinder skriftkulturen med de undervisningsmessige viktige forholdene, som elevenes motivasjon og deres arbeidsstrategier, mens ytringer blir til. I dette ligger at lese måten av figur 1 endrer seg når den ytre sirkelen kommer til. I det skriftkulturelle perspektivet som jeg startet med, så vi at samspillet mellom de to indre sirklene sto i fokus. I det utdanningsmessige perspektivet vi nå har kommet fram til, ligger det imidlertid en implisitt invitasjon til å lese modellen fra den ytterste sirkelen og innover. Hva dette innebærer, skal jeg omtale i neste avsnitt.

Semiotiske verktøy for ulike skrivehandlinger

Avsnittet ovenfor kan gi inntrykk av at bare relasjonen mellom situasjonsbestemte skriveformål og kulturelle grunnfunksjoner er sentrale for et gyldig skrivebegrep, men slik er det ikke. Enhver tekst, også elevytringer, trenger ulike slags «tegn» – semiotiske verktøy – for å kunne formuleres. I figur 2 nedenfor viser jeg relasjonen mellom elevers skrivehandlinger (formålsorienterte ytringer) og deres bruk av semiotiske verktøy for å utføre disse handlingene.

I denne figuren er de skriftlige uttrykksmidlene delt inn i fire hovedområder. Det første av disse omhandler bruken av ulike tegnsystemer («modaliteter»). I vår del av verden er vi oppdratt i en «verbosentrisk» kultur, hvor utdanningssystemet tilstreber en skriving som kun benytter



Figur 2

et verbalt (grafologisk) tegnsystem, mens våre små barn dokumenterbart lærer seg å skrive i en multimodal verden hvor de bruker både tegning og bokstaver (Kress, 1994). Den tradisjonelle undervisningen fører på dette området derfor til en ironisk avlæring. Hopper vi til de voksnes verden utenfor skolen, ser vi at multimedia blir stadig viktigere, og grafiske uttrykksmidler er en essensiell del av den profesjonelle skrivingas profesjonalitet. En gyldig skriveopplæring må følgelig dreie seg om mer enn skriftegn i snever forstand, og nye læreplaner bygger på en anerkjennelse av dette. Dermed må også en gyldig vurdering ha et rikere perspektiv på elevers bruk av de ulike skriftlige modalitetene enn før.

Det andre hovedområdet dreier seg om ulike former for strukturering. Skrijving er en sekvensiell aktivitet, men hvordan plasseres ulike elementer i en skriftlig ytring sekvensielt i forhold til hverandre? Og «bolkes» innhold i noen form for hierarki, som for eksempel avsnitt eller seksjoner? Skapes det videre noen funksjonell, logisk eller språklig sammenheng mellom de ulike sekvensielle eller hierarkiske elementene? Alle disse aspektene ved strukturering er relevante for kompetanseutvikling så vel som for vurdering av skrivekompetanse.

Det tredje hovedområdet under kategorien semiotiske verktøy gjelder manuelle aspekter ved skriving, som hvilke materielle skriveredskaper en bruker (blyant, penn, tastatur, mus o.l.) og skrivernes ferdighet i bruken av dem. En kan fristes til å tro at manuelle aspekter blir mindre relevante etter hvert som skriving i økende grad foregår med bruk av elektroniske medier. Slik er det imidlertid ikke. Tastaturferdighet vektlegger riktignok andre motoriske aspekter enn manuell skriftforming, men er både en like avgjørende forutsetning og et like risikofylt aspekt ved produksjon av tekst som den manuelle skriftformingen. Fenomenet «musearm» er bare ett eksempel på dette siste.

Det fjerde området fanger de mest tradisjonelt vektlagte aspektene ved skrevne tekster, nemlig hvilket ordforråd de inkluderer, og hvilke grammatiske strukturer som blir brukt. Også dette er sider ved bruken av semiotiske verktøy som konvensjonelt står sentralt i en vurderings-sammenheng. Dette tradisjonelle hovedfokuset er grunnen til at jeg har valgt å omtale denne gruppa sist. Når vi plasserer dette hovedområdet i en større semiotisk sammenheng, slik jeg gjør ovenfor, ser vi at perspektivet bør være bredere også når vi analyserer eller vurderer bruk av semiotiske verktøy i skriftlig kommunikasjon eller dokumentasjon. De er ikke bare formelle sider ved tekst, men i like stor grad ressurser for meningsskaping som andre ressurser. Og som ressurser for meningsskaping preges de av hensikten med skrivinga.

Gjennom figur 2 ser vi altså at vi kan lese semiotiske ressurser i lys av situerte ytringer eller skrivehandlinger (dvs. utenfra og innover). En slik lese måte har vesentlig betydning for vurdering. Ulike skrivehandlinger medfører at de enkelte semiotiske ressursene får ulik relevans og vekt. Skriver vi en handlelapp eller en annen huskeliste, er verken tekststruktur eller tegnsetting av vesentlig betydning for om skrivehandlingen er vellykket eller ikke, men argumenterer vi for et kontroversielt standpunkt, får disse verktøyene en mer sentral rolle. Tilsvarende gjelder for bruk av litterære virkemidler. Skriver vi for å glede, underholde eller dikte opp en fiktiv tekstverden, vil slike virkemidler være mer relevante enn om vi skriver for å forklare et faglig fenomen.

Kompetansenivåer i et lærings- og utviklingsperspektiv

Skrivehandlinger og semiotiske ressurser er begreper som gir en avgjørende bakgrunn for å kunne vurdere elevers skrivekompetanse. Likevel mangler fremdeles den viktigste komponenten, en presist definert kompetansekomponeent. En skrivemodell som den jeg har presentert så langt, har nemlig ikke noe lærings- eller utviklingsperspektiv. For at den skal kunne gi et grunnlag for gyldig vurdering i skolen, er det behov for en presisering av rimelige kompetansenivåer på ulike alders- eller årstrinn. Det betinger et samspill hvor en elevtekst kan leses i lys av både spesifikke kompetansemål og en spesifikk vurderingsskala. Jeg skal nedenfor omtale disse etter tur.

Kompetansenivåer kan defineres på ulike måter. IEA-undersøkelsen (Gorman, Purves & Degenhart, 1988) benyttet en strategi hvor bredt definerte kompetansenivåer ble illustrert med mønstertekster for hvert nivå. Norske skolemyndigheter har valgt en annen tilnæringsmåte i sine læreplaner: Det defineres kompetansemål for hvert trinn i utdanningssystemet (årstrinn eller grupper av årstrinn). For utvalgsprøver på 5. og 8. årstrinn innebærer en slik tilnæringsmåte at en må definere kompetansemål for nettopp disse trinnene.

Det er en krevende øvelse å definere slike kompetansemål. Enhver som prøver å bruke læreplanen (LK06) som grunnlag for vurdering av konkrete tekster, vil oppdage at målene ikke er formulert på en tydelig eller spesifikk måte. Det er heller ikke god nok progresjon mellom målene for ulike trinn. Ekspertgruppa bak 2005-prøvene forsøkte derfor å sammenholde læreplanmål med empiriske studier av kompetanseutvikling på ulike årstrinn. Dette resulterte i en tentativt klassifisert oversikt over «trinnmål» for de da aktuelle årstrinnene (4, 7, 10 og 11). Arbeidet med å utvikle en slik oversikt viste imidlertid raskt at vi i betydelig grad mangler det empiriske grunnlaget som den bør bygge på. Derfor har en gruppe forskere (Matre mfl., 2009) utviklet et forskningsprosjekt hvor en i samarbeid med et panel av utvalgte lærere kan utvikle mer presise og vurderbare mål.

Mens IEA-undersøkelsen som nevnt benyttet en strategi med bredt definerte mål, valgte Matres gruppe en strategi med langt mer presiserte målformuleringer. Bakgrunnen for dette er nyere internasjonale studier som er publisert i løpet av de senere årene (f.eks. Doherty & Hilberg, 2007; Quality Counts, 2006). De viser at ikke bare blir vurdering mer

pålitelig med bedre presiserte læringsmål, men elevenes læringsutbytte blir også bedre. Særlig ser det ut til å gjelde for faglig svake elever eller elever med en sosial bakgrunn som ikke favoriserer dem i utdannings-systemet.

IEA-studien viste at mål- eller nivåformuleringer må følges opp av gode eksempler på elevtekster som viser at et mål eller nivå er oppnådd for at vurderingene skal bli pålitelige. Matre m fl (2009) velger å benytte samme strategi. Utvalgte kompetansemål vil bli illustrert med henvisning til kommenterte utsnitt fra elevtekster, og vil bli publisert på det nasjonale Skrivesenterets kommende nettside for lærere (www.skriving.no). Hvordan kan dette se ut i praksis? La meg illustrere med et tidligere upublisert eksempel fra 2005-prøvenes bedømmeropplæring. Eksemplet viser hvordan et læringsmål kan nås innenfor to ulike skrivehandlinger.

I et slikt oppsett blir først målet gjengitt. Deretter blir det begrunnet ut fra det aktuelle årstrinnet (her 7. trinn) og kommentert på en måte som viser hvordan det er ment å skulle forstås. Så kommer utklipp fra elevtekster som illustrerer hva måloppnåelse innebærer, også i tekster som kan ha svakheter på andre områder. Denne skjerpningen av blikket i retning av *ulike enkeltmål samtidig* er nødvendig for å forhindre at den som skal bedømme, danner seg et helhetsinntrykk som bygger på tilfeldige enkeltområder. En elev vil altså vanligvis ha nådd en rekke mål selv om elevens ytring har mange svakheter. En slik presisering innebærer videre at prøver kan bli et instrument i opplæring, hvor elever får informasjon om det de får til, samtidig som det alltid er områder hvor de må arbeide for å forbedre sin kompetanse.

Et skarpt bedømmerblikk på målformuleringer må også innebære at visse mål er mer relevante enn andre når en skal vurdere spesifikke skrivehandlinger. Det innebærer at en rent intuitiv («holistisk») vurderingsstrategi ikke samsvarer med det skrivebegrepet som er definert i figur 1 og 2. Et intuitivt hovedinntrykk kan lett undervurdere betydningen av hvilken skrivehandling en gitt elevytring utfører. Tilsvarende vil det være dårlig samsvar mellom skrivebegrepet og en rent analytisk vurderingsstrategi, hvor en først bedømmer detaljer hver for seg og så summerer opp alle disse i en totalvurdering. Også en slik strategi vil lett undervurdere betydningen av hvilken skrivehandling som faktisk blir utført. Det skrivebegrepet i figur 1 og 2 innebærer, er faktisk en «primærtrekksrelatert» vurderingsstrategi, hvor et grunnleggende primærtrekk (her: skrivehandlingen) legger sterke premisser for hvordan ethvert aspekt ved en

Variasjon i språkbruk	Sjuende årstrinn
<p><i>Mål:</i> Ha noe variasjon i språkbruk (setningsformer, ta ordklasser som adjektiver og adverb i bruk, bruke noe spesialisert ordforråd som er tilpasset oppgaven)</p>	
<p>I sjuende årstrinn er elever kommet så langt i sin utvikling av skrivekompetanse at de fleste grunnleggende delkompetanser er tilegnet, i det minste på et enkelt nivå. Særlig viktig er det i så måte at de fleste av dem er i stand til å skrive lengre tekster. Dette stiller andre krav til framstillingsform, bl.a. for at tekster ikke skal bli for monotone å lese.</p> <p>I sjuende årstrinn er de fleste elever i stand til å skape slik variasjon på flere nivåer, og variasjon vil komme til syne i flere sjangrer.</p> <p>Hos én elev kan det være et kunstnerisk prosjekt som fører til innslag av beskrivende elementer. Disse elementene kan i sin tur føre til variasjon også i setningsform:</p> <p>«Jeg kom opp av havet som skummet og bruste. Fant en sti som førte meg inn i en tett, tett bjørkeskog. Jeg var helt alene i en mørk skog. Jeg gikk sakte innover i skogen. Det var så vidunderlig, så uvirkelig! Stammene var så glatte å ta på...»</p> <p>At ordforråd kan tilpasses innhold og skrivekontekst på en annen måte enn ovenfor, går også fram av det sakprega eksemplet nedenfor. I en rapport fra leirskoleopphold skriver en annen elev:</p> <p>«Etterpå gikk vi bortover et ras, guttene kastet steiner nedover raset. Vi gikk oppover en morenerygg, det var slitsom å gå, så jeg tok ofte pauser. Vi fant krystallav, den var rosa og ligner på mange små krystaller satt til sammen. Nå skulle vi ned til en bekk. Alle sammen sprang ned og begynte å drikke. Det var snø og is rundt sidene av bekken. Bekken lå i en V-dal.»</p> <p>I dette utsnittet bruker en skrivesvak elev flere fagtermer som passer til turprosjektets naturfaglige fokus (ras, morenerygg, krystallav, V-dal).</p>	

elevytring bør bedømmes. Thygesen mfl. (2007) valgte derfor en primærtreksrelatert strategi for 2005-prøvene og fant ingen grunn til å gå bort fra denne i sin kritiske analyse av resultatene og erfaringene fra prøvene i etterkant.

Vurderingsskala

Annoterte kompetansemål gjør at vi omsider begynner å nærme oss noe som kan være vurderbart, men de er i seg selv ikke en tilstrekkelig forutsetning for dette. For å kunne vurdere må vi ha en definert målestokk å vurdere etter – en vurderingsskala. Spørsmål om vurderingsskala blir gjerne drøftet som et spørsmål om pålitelighet (se Berge, 2009 om tolkningsfellesskap), men har også en viktig og underkommunisert validitetsdimensjon: Er det godt samsvar mellom vurderingsskalaen og den underliggende kvalitetsnormen som kompetansemålene gir uttrykk for? Et eksempel kan klargjøre denne sammenhengen.

Da forskergruppa bak 2005-prøvene skulle velge vurderingsskala, opplevde de et dilemma. To av de årstrinnene som prøvene omfattet, var på grunnskolens barne- og mellomtrinn, det vil si trinn hvor karakterer ikke skal benyttes i norsk skole. Samtidig var det liten erfaring med noen som helst formell vurdering på disse trinnene. Det var derfor viktig for gruppa å unngå at prøvene *de facto* skulle medvirke til at karakterer eller en karakterlignende skala ble innført. Hva var i så fall alternativet? Evalueringen av Reform 97 hadde avdekket at «tilpasset opplæring» fremdeles var en kjerneutfordring i norsk grunnopplæring (Haug, 2003). Gruppa valgte å ta utgangspunkt i denne utfordringen og samtidig spille på læreres tause kunnskap om hva en med erfaringsmessig rimelighet kan forvente av elever på bestemte årstrinn. Utgangspunktet for skalaen var et midtre skalatrinn der en elevtekst er «omtrent som en kan forvente» på et utvalg av vurderingsdimensjoner. Samtidig vil noen elever være så langt over eller under denne forventningen at det vil være behov for tiltak ut over ordinær undervisning. Skalaen fikk derfor to ytterpunkter – *mye over* hva en kan forvente, og *mye under* hva en kan forvente. Denne tretrinns skalaen ble brukt i 2005-prøvene.

Resultatene viste imidlertid at skalaen ikke fungerte etter intensjonen (Thygesen mfl., 2007). Dels var det lav pålitelighet i bruken av den, og lærere følte at den ga lite presis informasjon ut over det de visste fra før.

Men mer alvorlig var det kanskje at skalaen neppe var i samsvar med den måten som lærere rangerer kompetanse på. Og dette var i så fall et validitetsproblem. Hvordan kan vi så forstå dette validitetsproblemet? I sin doktorgradsavhandling undersøkte Evensen (1986) hva elever og lærere opplever som problemer i de to skolefagene norsk og engelsk. Forarbeidet til studien viste at lærere typisk omtaler en lang rekke fenomener langs femdelte rangeringsskalaer av typen *meget lett – ganske lett – middels – ganske vanskelig – meget vanskelig* (og tilsvarende *altfor lite – altfor mye* eller *altfor ofte – altfor sjelden*, osv.). Det fins altså en innarbeidet norsk skolediskurs for å vurdere. Den tradisjonelle femdelte karakterskalaen som gikk fra *lite godt* til *svært godt* var i samsvar med denne diskursen, mens skalaen i de nasjonale prøvene sannsynligvis ikke var det. Dermed kunne det se ut til at de lærerne som bedømte tekstene fra de nasjonale skriveprøvene, fremdeles vurderte med utgangspunkt i en skolediskurs hvor den gamle karakterskalaen var en ikke-uttalt premis, selv etter at den offisielt er avskaffet og selv på årstrinn hvor karakterer ikke er i offisiell bruk.

Dette taler for at en femdelt rangeringsskala (en «ordinal» skala) kan være det som best treffer læreres underliggende vurderingspremiss, og dermed være valid for det vurderingsskjønnet som en gyldig prøve vil måtte kommunisere med. Det fins imidlertid ulike former for femdelte skalaer, og den gamle karakterskalaen var bare én av disse. Denne er det da også fortsatt uaktuelt å bruke av både juridiske og etiske grunner. Hva kan så alternativet være? Skolekritikken mot den tredelte skalaen fra 2005-prøvene rettet seg først og fremst mot den brede midtkategorien. Da ekspertgruppa bak prøven drøftet resultatene med tanke på videreføring, foreslo de derfor å underdele midtkategorien i tre – dvs. i en øvre, en midtre og en nedre del (Thygesen mfl., 2007). Slik ville den tredelte skalaen oversettes til en femdelt skala.

Det er imidlertid et åpent spørsmål hvorvidt en slik underdeling av en normalkategori gir en entydig rangeringsskala, eller om det i realiteten er en todelt skala vi fortsatt snakker om (dvs. *som* forventet eller *ikke som* forventet) med underdelinger av hver del. I så fall er det fortsatt et gyldighetsproblem med skalaen. Et alternativ kan være å etablere en mangedelt «masterskala» for hele utdanningsløpet, hvor det til ethvert årstrinn defineres fem aktuelle trinn. En slik skala vil entydig være en rangeringsskala, og den kan lettere vise at de beste elevene på ett trinn kan prestere bedre enn svakere elever på høyere trinn. Det gjenstår imid-

lertid å undersøke hvor godt en slik alternativ skala vil kommunisere med læreres vurderingsskjønn.

Krav til gyldige utvalgsprøver

Med framstillingen ovenfor er vi kommet til et punkt i min artikkel hvor det er mulig å konkretisere hvilke egenskaper de kommende utvalgsprøvene bør ha for å tilfredsstille de gyldighetskravene som jeg drøfter. Den mest grunnleggende egenskapen en utvalgsprøve bør ha, er at den bygger på et gyldig skrivebegrep, slik dette begrepet er forsøkt definert i figur 1 og 2 og omtalen av disse. Det var ingen kritikk mot denne skriveforståelsen i forbindelse med de nasjonale 2005-prøvene, og jeg ser ingen grunn til at norske myndigheter eller fagpersoner skulle ha sterke motforestillinger på dette punktet.

Kravet til innholdsgyldighet er imidlertid langt mer kontroversielt. Både myndigheter og utdanningssystem ønsker en prøve som innebærer minst mulig merbelastning for alle involverte parter. En enkel måte å minske prøvebelastningen på er å gjøre den faglig smal, slik at bare ett aspekt eller delaspekt blir prøvd. 2005-prøvene ble balansert med tanke på denne problematikken, ved at elevene på hvert årstrinn skrev to ytringer innenfor ulike «skrivehandlinger» (jf. den ytre sirkelen i figur 1). Selv denne innsnevringen fritok imidlertid ikke prøvene fra kritikk for at de var tidkrevende å gjennomføre.

En alternativ strategi kan være at hver elev kun prøves i én skrivehandling, mens denne varieres fra år til år, slik at en etter et visst antall år vender tilbake til den samme skrivehandlingen og dermed får holdpunkter for å gjennomføre trendstudier for eventuell kompetanseutvikling på nasjonalt eller regionalt nivå. Faren ved en slik strategi er imidlertid at de først valgte skrivehandlingene lett kan oppfattes som å være de nasjonalt sett mest sentrale for de aktuelle årstrinnene. Derfor vil det i så fall være behov for å supplere en smal flerårsmodell med et verktøy som kan ivareta bredden for alle som blir valgt ut til å delta i et gitt års prøve.

Gruppen bak 2005-prøvene foreslo å supplere prøvene med en «elevens tekstsamling». En slik tekstsamling var tenkt som en elevmappe («portofolio») hvor eleven samler ikke bare ulike skrivehandlinger, men også eksempler på utkast, respons og bearbejdede tekster. Videre var tanken at samlingen kunne innholde resultater av lokalt prosjektarbeid, slik at prøvene ikke skulle fortrenge lokale behov og lokalt utviklingsarbeid.

Dersom en smal flerårsstrategi skulle bli valgt for de kommende utvalgsprøvene, er smalheten i en slik strategi et enda sterkere argument for å supplere de nye prøvene med en tekstsamling som kan ivareta hensynet til innholdsvaliditet. For skolene vil en slik samling dessuten kunne fungere som en kjærkommen bro mellom prøvene og den øvrige undervisningen.

En av de største faglige utfordringene for kommende prøver er å basere dem på et empirisk grunnlag for presise og vurderbare målformuleringer. Det innebærer at prøveutvikling og forskning bør gå hånd i hånd, i tett test mulig samarbeid med et utvalg skoler. Også et dekkende utvalg av annoterte elevtekster vil være et faglig krevende og tidsmessig omfattende arbeid. Ironisk nok kan disse utfordringene leses som et argument for å velge nettopp en flerårsstrategi for prøvene. En slik strategi ville innebære at en kan begynne med nasjonale trendstudier først etter at prøvene har vært gjennomført i flere år. En så lang tidshorisont vil gi noe tid til nødvendig forsknings- og utviklingsarbeid.

Etter mitt syn blir det i kritikken mot 2005-prøvene sagt at en kriterie-relatert vurderingsstrategi bør videreføres, mens nye utvalgsprøver ikke kan benytte samme vurderingsskala. Som vi har sett, er det imidlertid et krevende spørsmål å definere alternativet. Også her bør prøveutvikling følges av et parallelt empirisk arbeid, slik at vi kan finne ut hvilket alternativ som best kommuniserer med læreres underliggende vurderingsskjønn.

Sammenfatning

I en tidligere artikkel har jeg illustrert kompleksiteten ved vurdering gjennom å vise hvor mange ulike typer faktorer som samspiller med de rent faglige (Evensen, 2009). Denne kompleksiteten er én grunn til at mange lærere opplever vurdering som den mest ubehagelige siden ved arbeidet sitt. I denne artikkelen har jeg gått dypere inn i bare én av de rent faglige faktorene, nemlig gyldighetsproblematikken. Som vi har sett, er vurdering et krevende felt, selv med en smalere tilnærming. Vi slipper ikke unna kompleksitetsutfordringen.

Hvis vi tar gyldighetsproblematikken på alvor, innebærer det at vi må gå dypt inn i våre forutsetninger for praktisk utdanningsarbeid. Begrepsdefinering er et teoretisk spørsmål som kan innebære et langvarig intellektuelt arbeid, slik også arbeidet i ekspertgruppa bak 2005-prøvene viser

(Thygesen mfl., 2007). Prøveutvikling er altså på ingen måte et rent praktisk anliggende. Tvert imot betinger det et samspill mellom teoretisk og praktisk arbeid, og det er i nettopp dette samspillet at det spennende ved feltet ligger.

Litteratur

- Bazerman, C. (2000). Letters and the social grounding of differentiated genres. I D. Barton og N. Hall (red.), *Letter writing as social practice* (s. 15–30). Amsterdam: John Benjamins.
- Berge, K.L. (2009). Er tolkningsfellesskap mulig å oppnå i skriveprøver? I O.K. Haugaløkken, L.S. Evensen, F. Hertzberg og H. Otnes (red.), *Tekstvurdering som didaktisk utfordring* (s. 44–54). Oslo: Universitetsforlaget.
- Cizek, G.J., Bunch, M.B. og Koons, H. (2004). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, 25, 31–50.
- Doherty, R.W. og Hilberg, R.S. (2007). Standards for effective pedagogy, classroom organization, English proficiency, and student achievement. *The Journal of Educational Research*, 101 (1), 24–34.
- Evensen, L.S. (1986). *Den vet best hvor sko(l)en trykker...* Upublisert doktorgradsavhandling, AVH, Trondheim.
- Evensen, L.S. (1997). Betydningen av å arbeide i et langsomt medium. I *Svenskläraryörelsenes årsskrift 1997* (s. 24–33). Stockholm: Natur och kultur.
- Evensen, L.S. (2009). Vurdering av skrivekompetanse: En kompleks utfordring. I O.K. Haugaløkken, L.S. Evensen, F. Hertzberg og H. Otnes (red.), *Tekstvurdering som didaktisk utfordring* (s. 15–23). Oslo: Universitetsforlaget.
- Goody, J. og Watt, I. (1963). The consequences of literacy. *Comparative Studies in Society and History*, 5, 304–345.
- Gorman, T.P., Purves, A.C. og Degenhart, R.E. (red.) (1988). *The IEA study of written composition I: The international writing tasks and scoring scales*. Oxford: Pergamon Press.
- Haug, P. (2003). *Evaluering av Reform 97. Sluttrapport frå styret for evaluering av Reform 97*. Oslo: Norges forskningsråd.
- Kress, G. (1994). *Learning to write*. London: Routledge.
- Matre, S. mfl. (2009). *Developing national standards for the teaching and assessment of writing*. Upublisert prosjektbeskrivelse for søknad til Norges forskningsråd.
- Miller, C.R. (1984). Genre as social action. *Quarterly Journal of Speech*, 70, 151–167.
- Olson, D.R. (1994). *The world on paper*. Cambridge: Cambridge University Press.
- Purves, A.C. (red.) (1992). *The IEA study of written composition II: Education and performance in 14 countries*. Oxford: Pergamon Press.
- Scribner, S. & Cole, M. (1981). *The psychology of literacy*. Cambridge, MS: Harvard University Press.

- Swales, J. (1990). *Genre analysis*. New York: Cambridge University Press.
- Thygesen, R., Evensen, L.S., Berge, K.L., Fasting, R.B., Vagle, W. og Haanæs, I.R. (2007). *Nasjonale prøver i skriving som grunnleggende ferdighet. Sluttrapport*. Stavanger: Nasjonalt senter for leseopplæring og leseforskning, Universitetet i Stavanger.
- Quality Counts. (2006). National assessment of educational progress. *Education Week*, 23 (5), 33.