

Gustaf Bernhard Uno Skar

Norges teknisk-naturvitenskapelige universitet

Arne Johannes Aasen

Norges teknisk-naturvitenskapelige universitet

DOI: <http://dx.doi.org/10.5617/adno.6280>

Å måle skriving som grunnleggende ferdighet

Sammendrag

Denne artikkelen undersøker utviklinga av skriveprøver i Norge (i 2005 og 2009–2017). Formålet er å drøfte utfordringer forbundet med det å måle elevers skriving som grunnleggende ferdighet gjennom skriveprøver. I undersøkelsen tar vi utgangspunkt i flere av de endringene som ble gjort fra de nasjonale prøvene i skriving i 2005 til de læringsstøttende prøvene som ble utviklet fram mot 2017. Et viktig grunnlag for undersøkelsen er tekniske rapporter som ble skrevet av de gruppene som lagde prøvene.

Å måle elevenes skriveferdigheter har vært faglig komplekst og kostbart. Artikkelen gjør rede for hvordan prøvene ble utviklet, hva som har kjennetegnet dem og hvilke endringer som har skjedd med dem i løpet av denne perioden, og belyser derved denne kompleksiteten. Vi redegjør for dilemmaer som følger med skriveprøveutvikling, men peker også på mulige løsninger for å ivareta god prøve kvalitet.

Nøkkelord: skriveprøver, skriving, skriving som grunnleggende ferdighet, validitet, reliabilitet

Measuring writing as a key competence

Abstract

This article investigates the development of national writing tests in Norway (in 2005 and 2009–2017). The purpose of the study is to present and discuss challenges associated with the assessment of writing proficiency through standardized writing tests. The article focuses on changes that were made between the first administration of national writing tests in 2005 and the subsequent administrations following a new test design in 2009. Key sources that document this change are technical reports from the test developers.

Assessing writing through national writing tests has proved to be complex, posing several challenges including numerous design choices, and has also proved to be expensive. To shed light on these complexities, this article provides

information on how the tests were developed, how they can be characterized and what type of changes that were made in the period 2005–2017. The article explicates some dilemmas associated with development of writing tests, but also some viable solutions when aiming to produce high quality tests.

Keywords: writing assessment, writing, writing as a key competence, validity, reliability

Innledning

På begynnelsen av 2000-tallet ble skriving løftet fram som en av basiskompetansene i skolen. I 2006 ble skriving gjennom læreplanreformen Kunnskapsløftet (LK06) en av fem grunnleggende ferdigheter. Dette innebar en anerkjennelse av at skriving er et sentralt verktøy for læring og for vurdering av elevenes kompetanse i fagene i skolen. Med skriving som grunnleggende ferdighet fulgte et behov for lærere og samfunnet for øvrig for å kartlegge elevers skrivekompetanse.

Å måle skrivekompetanse er imidlertid vanskelig, og skriving er gjerne omtalt som problembarnet i vurderingsklassa. En årsak til dette er at skrivekompetanse er et komplekst kulturelt og kognitivt fenomen som ikke kan måles gjennom enkeltoppgaver. Andre årsaker har å gjøre med målingas instabilitet: elever presterer svært ulikt når de skriver forskjellige oppgaver, og det er krevende å oppnå akseptabelt samsvar mellom lærere som vurderer elevers skriving gjennom elevtekster (Bouwer, Béguin, Sanders & van den Bergh, 2015).

Trass i vanskelighetene med å måle skriving som grunnleggende ferdighet har det blitt gjort to forsøk med gjennomføring av skriveprøver¹, henholdsvis nasjonale prøver i skriving, og nasjonale utvalgsprøver og læringsstøttende prøver i skriving. Nasjonale prøver i skriving ble gjennomført i 2005, men ble stoppet allerede i 2006, nettopp på grunn av problemer med å oppnå akseptabelt samsvar mellom lærerne som vurderte prøvene. I 2009 begynte arbeidet med å utvikle læringsstøttende prøver i skriving. Disse prøvene ble stoppet i 2016, men denne gangen av budsjettmessige årsaker.

Formålet med denne artikkelen er å løfte fram og samle noen av de sentrale innsiktene fra arbeidet med utvikling av nasjonale og læringsstøttende prøver i skriving. Vi gjør rede for hvordan prøvene ble utviklet, hva som har kjenne-tegnet dem, og hvilke endringer som har skjedd med dem i utviklingsperioden 2003–2016. Dette løfter fram hvilke utfordringer som har vært forbundet med

¹ Tre prøver er omtalt i teksten. De nasjonale prøvene i skriving og utvalgsprøvene i skriving er prøver som er gjennomført av elever. De læringsstøttende prøvene er prøver som er utviklet på grunnlag av utvalgsprøvene i skriving, og lærere kan frivillig bruke de læringsstøttende prøvene som en ressurs i skriveopplæringa uten at resultatene blir innrapportert.

det å måle elevers skrivning som grunnleggende ferdighet gjennom skriveprøver. Forskingsspørsmålene er derfor:

- Hvilke sentrale utfordringer var forbundet med utviklinga av skriveprøver?
- Hvilke forandringer gjennomgikk skriveprøvene for å møte disse utfordringene?

Artikkelen er bygd opp på følgende måte: Først greier vi ut om skrivning og om skrivning som grunnleggende ferdighet. Deretter tegner vi en kort historikk over skriveprøver i Norge. Etter dette følger en gjennomgang av relevant litteratur om skrivevurdering gjennom en modell som gir oversikt over sentrale aspekter ved vurdering av skrivning. Noen elementer fra modellen blir brukt i den resterende delen av artikkelen, som presenterer de fire dimensjonene ved skriveprøvene som gjennomgikk de største forandringene. Ei kort drøfting avslutter presentasjonen av hver dimensjon.

Skriving

Skriving kan med utgangspunkt i Bereiter og Scardamalia (1987) forstås som det å komponere tekster som skal leses av noen som ikke er til stede i skriveøyeblikket. Skrivning kan også være komposisjon av en tekst som man selv leser i framtida.

Skrivekompetanse kan forstås som det å ha kapasitet til å skape tekster som realiserer ulike formål som er kulturelt og sosialt situerte (f.eks. Evensen, 2010). Skrivning er ikke en naturlig aktivitet, og utvikling av skrivekompetanse tar opp til 20 år (Kellogg, 2008) og foregår i forskjellige stadier (Bereiter & Scardamalia, 1987). Med hjelp av skriveopplæring vil et barn gjerne gå fra å kunne forme bokstaver og formulere enkle og korte tekster til å kunne lage tekster basert på refleksjoner over hvordan en kan utforme ideer i skrift med tanke på mottaker og relevans for skrivesituasjonen (Rijlaarsdam et al., 2012, s. 196f). Skrivning består av mangefasetterte kognitive prosesser (Flower & Hayes, 1981), og i takt med at skriverne utvikles, lærer de med ordene til Rijlaarsdam et al. (2012, egen oversettelse) å «bruke mentale ressurser (oppmerksomhet, minne og prosessering) i vekselspillet mellom ideer og intensjoner på den ene siden og ulike formuleringer og tekstutforminger på den andre siden». Disse prosessene krever modning, kognitiv utvikling og ikke minst erfaringer med skrivning.

Applebee (sitert i Jeffrey, Elf, Skar & Wilcox, u.u.) mente at skriveutvikling ikke er mulig å skille fra sosialt konstruerte måter å kommunisere på:

Once we accept that writing is a socially constructed set of ways of communicating, the notion of writing development becomes inescapably intertwined with notions of curriculum [no. undervisning]. Students will learn those genres, skills, and strategies with which they are given experience through their school, and are much less likely to

learn those that the schools ignore, reject, or simply postpone for attention in later years.

Implikasjonene av resonnementet til Applebee er altså at skriveutvikling er avhengig av skriveopplæring og de skriveerfaringer elever får gjøre gjennom utdanningsløpet. Skriveutvikling er med ordene til Camp (2012, s. 101) noe som henger sammen med verdier, sosiale interaksjoner og kulturelle konvensjoner og derfor er kontekstuell heller enn universell. En annen implikasjon av dette resonnementet er at skriving som «problembarnet i vurderingsklassen» ikke bare har med måletekniske utfordringer å gjøre, men også med utfordringer med å lage oppgaver som på en gyldig måte tar hensyn til nettopp kontekstuelle faktorer ved skriveutvikling.

Det finnes mange teoretiske modeller som utgår fra skrivingas og skriveutviklingas kulturelle aspekter. Disse modellene kan for eksempel fokusere på skrivers kommunikative intensjoner innenfor en gitt situasjons- og kulturkontekst (Berge, Evensen & Thygesen, 2016), eller fokusere på hva som utmerker tekster som oppfyller kulturens forventinger (Rose, 2016). En tredje modellering er nasjonale eller regionale standarder nedfelte i læreplaner eller andre styringsdokumenter for skolen (Bazerman et al., 2017). Slike standarder beskriver implisitt hva skriving og tekstkvalitet er, gjennom å angi hva en skriver forventes å kunne gjøre med skrift. Standardene beskriver dessuten eksplisitt en tenkt progresjon i skriveutviklinga.

Vurdering av skriving i Norge – en kort bakgrunn

I begynnelsen av 2000-tallet ble det gjennomført omfattende utredningsarbeid om utdanningssystemet og innholdet i opplæringa i skolen. Arbeidet resulterte i flere NOU'er, stortingsmeldinger og lovforslag (f.eks. St.meld. nr. 30 (2003–2004)). Tre av de viktigste resultatene i vår sammenheng var 1) innføring av nasjonale prøver, 2) innføring av nye læreplaner (Kunnskapsløftet), og 3) innføring av grunnleggende ferdigheter.

Hensikten med nasjonale prøver var å gi politisk og byråkratisk ledelse i utdanningssektoren informasjon om elevenes læring og utvikling av grunnleggende ferdigheter, men også å gi foreldre og elever «bedre grunnlag for å stille krav og/eller medvirke til forbedringer i opplæringa» (St.prp. nr. 1, Tillegg nr. 3 (2002–2003)). Prøvene skulle gi en indikasjon på om opplæringa la til rette for elevenes utvikling av grunnleggende ferdigheter, altså at opplæringa skulle bli tilpasset den enkelte elevens behov (NOU 2002, s. 10).

I Norge er det hovedsakelig norskfaget som har og har hatt ansvaret for måling av elevers skrivekompetanse. Den viktigste målinga har vært eksamen i norsk, som har vært den offisielle operasjonaliseringa av læreplanen i norsk. Norskeksamen har blitt gjennomført i flere ulike format (Anderson, 2015, s. 40;

Berge, 2015; Norendal, 2016). I sitt seneste format består eksamen av et teksthefte og et batteri med oppgaver der elevene skal besvare noen av dem. Oppgavene ber som regel elevene om å ta utgangspunkt i tekstheftet for å utføre ei skrivehandling, og oppgavene kan kategoriseres som integrerte (dvs. måling av lesing, skriving og norskfaglig kunnskap), noe som er rimelig med tanke på at prøva ikke bare skal måle skrivekompetanse, men en helhetlig norskfaglig kompetanse.

Det har blitt gjennomført få validitetsstudier av norskeksamen, men noen studier har undersøkt reliabiliteten (f.eks. Berge, 1996, 2005b; Krogh, 2016), og stort sett har denne vært problematisk lav, selv om Berge (2005b, s. 109) konkluderte med at nivåene var rimelige for visse teksttyper i avgangsprøva i norsk hovedmål. Andre mer kvalitativt innretta undersøkelser har vist hvordan tekstlingvistiske trekk har sammenheng med karakter (f.eks. Skjelten, 2013), og oppgavene har blitt undersøkt ut fra et dannelsingsperspektiv (Norendal, 2016).

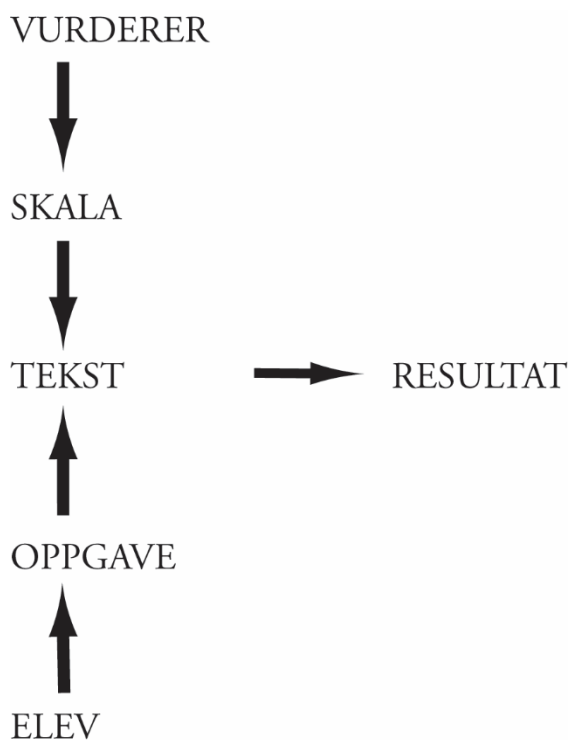
Ved siden av norskeksamen er det henholdsvis nasjonal prøve i skriving og utvalgsprøver i skriving som har undersøkt elevers skrivekompetanse i Norge. I 2005 var prøva ei «populasjonsprøve» der de fleste elevene på 4., 7. og 10. årstrinn deltok. Også elever i Vg1 skulle delta, men en boikott medførte at bare fire prosent av elevene gjennomførte prøva (Thygesen et al., 2007, s. 83).

Fra 2010 ble skriveprøva gjennomført som ei utvalgsprøve med to hovedformål. Det første formålet var å utvikle grunnlaget for læringsstøttende prøver i skriving. Utvalgsprøvene bidro til at det var mulig å utvikle oppgavesett, mestringsnivåbeskrivelser og eksempeltekster. De læringsstøttende prøvene skulle være, og er, prøver som er fritt tilgjengelige for lærere som vil få informasjon om egne elevers skriveferdigheter (se www.skrivesenteret.no). Det andre formålet med utvalgsprøvene var å gi informasjon om hvordan elever mestrer skriving som grunnleggende ferdighet.

Forsking på utvalgsprøvene i skriving har inkludert undersøkelser av de teoretiske utgangspunktene for konstruktdefinisjonen (Berge, 2005a; Evensen, 2010), prøvegjennomføring og tekniske kvaliteter (Fasting, Thygesen, Berge, Evensen & Vagle, 2009; Skar, 2017; Thygesen et al., 2007), spørsmål rundt tolkingsfellesskap (Jølle, 2014; Skar & Jølle, 2017) og betraktninger over politiske og byråkratiske prosesser rundt prøvene (f.eks. Berge, 2007).

Å måle skriving

Kompleksiteten i vurdering av skriving tydeliggjøres når vi tar i betraktning de interaksjonene og komponentene som inngår i typiske vurderingsprosesser. Følgende modell (McNamara, 1996; Skar & Aasen, 2016) viser hvordan en elevs skriving og en lærers vurdering av en elevtekst kan bli et skrivevurderingsresultat:



Figur 1. McNamaras (1996) modell for språkprøver, etter Skar og Aasen (2016).

I midten av modellen står teksten, produktet av skrivinga. Nederst i modellen står eleven som, for å skulle produsere en tekst, må fortolke ei skriveoppgave og gjennomføre skriveprosessen som har teksten som produkt. Oppgava er det avgjørende leddet mellom eleven og den teksten som blir skrevet. Øverst står vurdereren som vurderer teksten med hjelp av en eller annen form for skala, gjerne en vurderingsmatrise. Skalaene er avgjørende for vurdererens lesing av teksten og er det instrumentet vurdereren har for å sortere tekster i kvalitetsnivåer. I de kommende avsnittene ser vi nærmere på oppgaver, skalaer og vurderere/vurdering for å ytterligere belyse kompleksiteten som ligger i måling av skrivekompetanse. Vi sier imidlertid først noe om utgangspunktet for utvikling av oppgaver og skalaer, nemlig konstruktdefinisjonen.

Konstruktdefinisjon

Skrivehjulet (Berge et al., 2016) er en modell som har ligget til grunn for nasjonale prøver og læringsstøttende prøver i skriving i Norge. Den ligner på modeller som beskriver språk- og skrivekompetanse ut fra et funksjonelt eller kommunikativt perspektiv (f.eks. Bachman, 1990; Bachman & Palmer, 2010; Purpurra, 2010; Purves, Gorman & Takala, 1988). Felles for disse modellene er at de skiller mellom kunnskap om språk og kunnskap om å anvende språk. For eksempel identifiserer den ofte siterte modellen til Bachman for kommunikativ språkkompetanse (Bachman, 1990; Bachman & Palmer, 2010) to komponenter: språkkunnskap og strategisk kunnskap. Språkkunnskap deles i modellen opp i organisatorisk kunnskap, noe som her betyr å kunne anvende skrift for å skape en (forståelig) tekst, og pragmatisk kunnskap, noe som her viser til de

funksjonelle aspektene ved språket, for eksempel sjangerkunnskap. Den strategiske kunnskapen kommer til uttrykk i de valg en skriver tar i en gitt kommunikasjonssituasjon for å oppnå kommunikative mål med et relevant innhold.

Skriveoppgaver

Ei skriveoppgave skal ideelt sett være ei operasjonalisering av konstruktdefinisjonen. Problemet med denne forutsetninga er at konstruktdefinisjoner som fanger bredde og helhet i skrivekompetanse, vil kunne realiseres på utallige måter. Derfor blir det krevende å gi en generell definisjon på hva ei god skriveoppgave er for noe. Det finnes mange ulike typer oppgaver, og de kan være forskjellige på ulike måter. Noen eksempler på dimensjoner ved oppgaver som påvirker rammene for å skrive tekster, er:

- spesifisering av sjanger, teksttyper og skrivehandlinger. Disse aspektene relaterer til det å skrive i ulike situasjoner, f.eks. om ulike kunnskapsobjekter og ulike måter å utvikle og kommunisere kunnskap på. Dette medfører videre at ulike fag har ulike konvensjoner for skriftlig mediering, med andre ord ulike sjangre eller ulike måter å realisere samme sjanger på i ulike fag (Shanahan & Shanahan, 2008).
- retorisk spesifisering (Otnes, 2015), dvs. om oppgava konstruerer en kommunikasjonssituasjon eller ikke. Noen oppgaver angir mottaker og publiseringssted, og i visse tilfeller er dette autentiske sammenhenger med faktiske lesere, mens i andre tilfeller er det tale om fiktive sammenhenger der den faktiske mottakeren, vurdereren, ikke er nevnt som mottaker. Høg retorisk spesifisering krever skrivere som aksepterer en dobbel kontrakt; den ene med fiktive lesere (f.eks. lesere av ei avis) og den andre med de faktiske leserne (læreren, vurderere).
- vilkår for å løse oppgava som tid, sted, hjelpemiddel. Det er f.eks. forskjell på å skrive til skriveoppgaver under kraftig standardiserte forhold (tester, eksamen osv.) og skriveoppgaver der elevene får skrive hjemme.

Den viktigste dimensjonen er kanskje likevel innhold. Innholdsdimensjonen vil i høg grad bidra til mangfoldigheten i skriveoppgaver: Det er stor forskjell på å skrive til ei oppgave om prosesser i naturen versus ei oppgave om hva man har gjort i sommerferien. Det finnes studier som tilsier at forskjell i innhold har like stor påvirkning som forskjell i sjanger eller teksttype (Rijlaarsdam et al., 2012, s. 211); det kan for eksempel være like krevende å skrive beskrivende som det å skrive argumenterende om to forskjellige temaer. Dessuten er det slik at hvis eleven ikke har noe å skrive om, vil hun ikke kunne bruke sin kompetanse innenfor områder som tekstopbygging, syntaks og tegnsetting. Det at eleven i det hele tatt kan generere et innhold, vil ha mye å si for hva slags tekst eleven skriver, men også for kvaliteten på teksten (Rijlaarsdam et al., 2012).

Vurderingsskalaer

I skrivevurdering består skalaer av vurderingskriterier som beskriver nivåer av tekstkvalitet. Hvilke kriterier som inngår i skalaen, er bestemt av hvilke egenskaper ved skrivekompetansen man ønsker å måle. Ved såkalt analytisk vurdering har man flere skalaer, eller vurderingsområder, for ulike egenskaper ved tekstene. Ved såkalt holistisk vurdering er hvert kvalitetsnivå kopla til kriterier som beskriver teksten som helhet (Knoch, 2009; Weigle, 2002).

Vurdereren

Skrivekompetanse er ikke mulig å observere direkte, men krever at en leser vurderer kvaliteten til tekster for så å gjøre inferenser om denne kompetansen. Det finnes to hovedtyper av utfordringer forbundet med vurdering av elevtekster. For det første innebærer selve vurderinga en kompleks kognitiv prosess. Oppgava for en vurderer er å identifisere de kriteriene som best mulig beskriver elevteksten. Men når ei gruppe elever skriver tekster til ei oppgave, så vil de løse oppgava på svært ulike måter, altså vil elevtekstene se svært forskjellige ut. Det vil derfor finnes flere typer elevsvar som representeres av ett og samme kriterium. Dette krever god tekst- og lesekompetanse der vurdererne må fortolke elevtekstene i lys av sin forståelse av kriteriene.

En annen utfordring ved vurdering er at ulike vurderere vil bruke vurderingsverktøyene på ulike måter. Det finnes flere eksempler på dette (se Myford & Wolfe, 2003 for en uttømmende liste). Det kan være vanskelig for vurderere å skille ulike vurderingsområder, noe som for eksempel betyr at vurdereren i møte med en tekst med mange skrivefeil lar dette trekke farge vurderinga på andre områder. En annen utfordring kan være at ulike vurderere bruker skalaen ulikt; for eksempel vil noen foretrekke trinn på midten av skalaen mens andre i større grad bruker ytterdelene av skalaen. Et tredje eksempel på utfordring er at vurderere skiller seg i strenghet. Og til sist kan det være en utfordring at vurderere har ulike forestillinger om formålet med vurderinga, for eksempel om den skal bidra til å motivere eleven til revisjon eller om den skal gi en sann representasjon av elevtekstens kvaliteter (Jølle & Skar, u.u.).

Reliabilitet og validitet

En vanlig problematikk i vurdering av skrijving er hvorvidt reliabilitet eller validitet skal prioriteres. Reliabilitet kan defineres som konsistens i vurderinga, noe som innebærer at resultatet ikke varierer avhengig av for eksempel hvem som vurderer eller hvor/når prøva gjennomføres. Validitet innebærer at det er mulig å tolke testresultatet på en meningsfull måte i forhold til konstruktdefinisjonen, og dessuten at anvendelsen av testresultatet er rimelig i forhold det man har hatt som intensjon å måle (Kane, 2013). Hva som prioriteres vil få konsekvenser for hvordan skrivekompetanse måles. Enten kan man måle skrivekompetanse indirekte gjennom oppgaver der elevene for eksempel skal identifisere grammatiske feil (altså «interlinear exercises», Breland, Bridgeman

& Fowles, 1999), eller direkte gjennom oppgaver der elevene oppfordres til å skrive tekster. Valget mellom indirekte eller direkte måling av skrivekompetanse har ofte, forenklet sagt, vært et valg mellom å prioritere reliabilitet eller begrepsvaliditet (Huot & Perry, 2009); indirekte måling har den fordelen at den eliminerer vurderervariasjon, mens direkte måling har den fordelen at elevene får utføre de handlingene som man er interessert i å måle.

Flere validitetsforskere har vært sterkt kritiske til skillet mellom reliabilitet og validitet. Kane, Crooks og Cohen (1999) illustrerte i sin viktige artikkel om validering av såkalt «performance assessment» at det å kunne gjøre ekstrapoleringer fra et testresultat til et måldomene (f.eks. at resultatet på eksamen i norsk gir grunnlag for å trekke slutninger om elevenes norskfaglige kompetanse), ikke bare krever at oppgava speiler handlinger i måldomenet, men også krever generaliserbare resultat. I så fall skal resultatet bli det samme uavhengig av hvem som vurderer teksten eller hvilke oppgaver i et utvalg av lignende oppgaver som eleven har gjennomført. Derfor er reliabilitet en forutsetning for validitet.

Endringer i nasjonale prøver 2003–2016

I resten av artikkelen rettes fokuset mot endringene fra nasjonale prøver i skriving til utvalgsprøvene i skriving som la grunnlaget for utvikling av læringsstøttende prøver. Med utgangspunkt i modellen til McNamara gir tre dimensjoner ved vurdering av skriving inngangene til undersøkelsen: oppgaver, skalaer og vurdering av tekster. Vi innleder imidlertid med konstruktdefinisjonen. For hver dimensjon er framstillinga bygd opp slik at vi først presenterer aspekter ved prøva i 2005, deretter innslag fra den evalueringa som den eksterne gruppa med eksperter gjennomførte etter innføringa av nasjonale prøver, og til sist viser vi hvordan prøva ble endret fra 2005 til 2016. I forbindelse med hver dimensjon presenterer vi en drøfting. Framstillinga bygger til stor del på oppdragsbrevet fra Utdanningsdirektoratet (2014) og tekniske rapporter (Fasting, 2011; Skar, 2017; Skar, Evensen & Iversen, 2015; Skar & Iversen, 2016).

Fra Skrivekompetanshjulet til Skrivehjulet

Da de nasjonale prøvene skulle utvikles på begynnelsen av 2000-tallet, fantes det ingen «offisiell» norsk definisjon av skrivekompetanse, selv om norsk-eksamen indirekte definerte hva skriving er og hvordan skriving skulle vurderes. Innenfor andrespråksforskninga fantes flere definisjoner. Der hadde man siden 1980-tallet fokusert på kommunikative språkkompetanser, og man hadde gått fra modeller som fokuserte på språkets formside til modeller der det å kunne anvende språket kommunikativt kom mer i sentrum (Purpurra, 2010). Gruppa som lagde nasjonale prøver i skriving, mente at det var viktig med en modell som kunne beskrive funksjonell skriving, og derfor konstruerte de det såkalte

Skrivehjulet. I utgangspunktet i 2005 ble denne modellen kalt Skrivekompetansehjulet, og den bestod av:

- skrivekompetanser, f.eks. det å kunne overtale noen
- handlinger og formål, f.eks. det å kunne uttrykke meninger
- skriftlig mediering, dvs. det å skape tekster ved hjelp av tegn (f.eks. bokstaver og bilder)
- skriveprosesser, f.eks. revisjon av tekst

Ifølge Berge (2005a, s. 177) var skrivekompetanser det samme som «grunnleggende meningsskapingsoperasjoner», og skrivehandlinger og skriveformål var det samme som «bruken» av skriving. Et eksempel var skrivekompetansen «kunnskapslagring og strukturering» som var kopla til handlinga og formålet «bruke skrift som st[ø]tte for kognitive operasjoner». Hensikten var å lage et instrument som gjorde det mulig å måle hvordan en elev gjennomfører en grunnleggende funksjon, heller enn å måle elevens evne til å skrive innenfor rammene til bestemte sjangrer. Like fullt ble distinksjonene mellom kompetanser, handlinger, formål, skriftlig mediering og skriveprosesser vurdert å ikke være tydelige nok, noe som førte til at modellen ble kraftig revidert i perioden 2010–2012 (Dagsland, 2018).

I Berge et al. (2016) presenteres den seneste versjonen av modellen (se figur 2), som nå består av tre lag: et ytre som beskriver «skrivehandlinger», et mellomlag som beskriver «skriveformål» og et indre som beskriver «skrivingsverktøy». Skrivehandlingene som inngår i modellen er å *samhandle*, å *reflektere*, å *beskrive*, å *utforske*, å *forestille seg* og å *overbevise*. Disse kan enklest forstås som typiske tekster (se f.eks. Berge et al., 2016, s. 175 og s. 180), og opphavsmennene bak modellen viser for eksempel til arbeidstegning som en typisk manifestasjon av skrivehandlinga å beskrive. En skrivehandling er i sin tur et uttrykk for både intensjoner om å skape mening og kommunikative formål.

Drøfting av konstruktdefinisjonen. Et viktig poeng med Skrivehjulet er at måling av skriveferdighet forutsetter at en skriver testes i å skrive på en rekke forskjellige måter som er tilpasset ulike situasjoner og dermed ulike formål. Skrivehjulet er altså lagd for å fange både dybden (handling, formål og mediering) og bredden (*ulike* formål og skrivehandlinger) i skriveferdighet. Ulike skrivehandlinger resulterer i ulike tekster; det vil for eksempel være forskjeller mellom en tekst som skal *reflektere* over dialektvariasjon i Norge og en tekst som skal *argumentere* for dialektvariasjon.

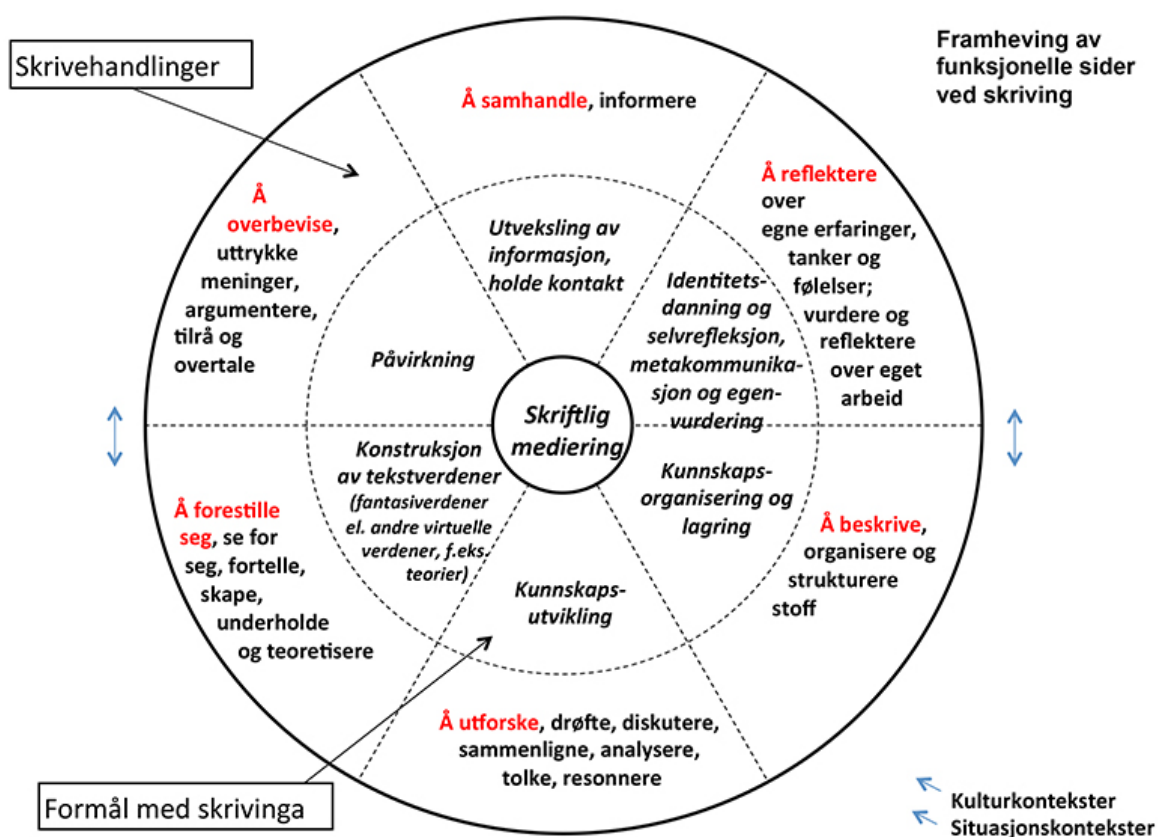
Til forskjell fra andre modeller beskriver ikke Skrivehjulet typiske trekk ved ulike tekster, og dette er en av styrkene til modellen; svært ulike tekster kan bli vurdert som gyldige svar på éi oppgave. På den annen side er dette også en av utfordringene til modellen; den presenterer ikke tydelige distinksjoner mellom skrivehandlinger, noe som fører til at det er vanskelig å hevde at en elev behersker ei skrivehandling og ikke ei anna. Videre er det åpenbart at modellen

ikke beskriver andre viktige aspekter ved skrivekompetanse, som kognitive prosesser, strategier, skriverens forestillinger om seg selv som skriver, og skriverens motivasjon og holdninger. Derfor var det rimelig å gå bort fra kompetansebegrepet i modellen og heller fokusere på handlinger og formål.

I den første evalueringa av de nasjonale prøvene i 2005 ble Skrivehjulet introdusert som en landvinning, både som teoretisk konstrukt og som pedagogisk potensial:

Den store styrken til skriveprøven er at den representerer et pedagogisk verktøy, utviklet ved hjelp av en grundig teoretisk analyse av hva skrivekompetanse består av. [...] Gjennom dette utviklingsarbeidet har man fått et begrepsapparat og et viktig verktøy i arbeidet med å heve vurderingskompetansen i skrijving i norsk skole. (Lie, Hopfenbeck, Ibsen & Turmo, 2005, s. 80)

Skrivehjulet fortsatte å være konstruktdefinisjonen for prøvene etter 2005.



Figur 2. Skrivehjulet. Fra Solheim og Matre (2014).

Fra integrerte oppgaver til prompt-baserte oppgaver

To viktige spørsmål ved utvikling av skriveoppgaver er: Hva skal elevene skrive om, og hvor skal stoffet de skriver om hentes ifra? I 2005 valgte man ei løsning med integrerte skriveoppgaver der alle elever i grunnskolen skulle skrive svar på to oppgaver med temaet «planeten Mars». Elevene fikk tilgang til informasjon om dette temaet gjennom et hefte med tekster som på ulikt vis tematiserte planeten Mars. Dessuten skulle elevene skrive to oppgaver der de gjennomførte

to forskjellige typer skrivehandlinger (se vedlegg 1). Den nye konstruktdefinisjonen bygde jo på en forståelse av at kompetanse i skriving ikke kan måles gjennom éi oppgave, men helst gjennom flere ulike skriveoppgaver som lar elevene gjennomføre ulike skrivehandlinger.

Som vi kan se i oppgavene i vedlegg 1, ble elevene bedt om å utføre to skrivehandlinger og skape tekster, og i motsetning til eksamen i norsk var ikke oppgavene sjangerbaserte. Dessuten inneholdt oppgavene en viss grad av retorisk spesifisering (Otnes, 2015), det vil si spesifisering av kommunikasjons-situasjon og mottakere. Dette var en forutsetning for at det skulle være mulig å vurdere om elevene kunne tilpasse tekstene sine til kommunikative sammenhenger.

Verken prøvekonstruktørene eller eksterne granskere uttrykte noen egentlig skepsis til oppgavene, men i evalueringa av prøvene ble det notert at den andre oppgava «ber om ikke mindre enn en refleksjon rundt et av de store spørsmål, meningen med livet, etikk og estetikk, hvem er jeg osv.» (Lie et al., 2005, s. 71).

Da arbeidet med skriveprøver startet opp igjen i 2010, var prøveformatet endret. For å rendyrke målinga av nettopp *skrivning* som grunnleggende ferdighet og redusere forberedelsesarbeidet, fikk Skrivesenteret i oppdrag å utvikle såkalte «impromptu-oppgaver» (Weigle, 2002). Disse innebar at elevene skrev tekster som svar på oppgaver som ikke forutsatte at de skulle lese andre tekster. Argumentene for denne forandringa var praktisk fundert og relatert til validitet; å minke forberedelsestid var positivt for det praktiske arbeidet med å gjennomføre prøva, og å utelukke andre ferdigheter (som lesing) hadde betydning for mulighetene for å argumentere for validiteten til prøveresultatene. Prøva skulle avdekke elevenes ferdigheter til å kommunisere med hjelp av skrift, ikke om de var gode lesere og så videre.

I det nye formatet skulle ei oppgave leses høgt for elevene og etterfølges av ei lærerledet 15 minutter lang muntlig idémyldring før elevene skulle skrive en tekst. Se vedlegg 2 for et eksempel på ei oppgave.

Oppgavene etter 2010 bestod av en prompt, altså et skriveoppdrag, som var basert på ei av skrivehandlingene i Skrivehjulet. Den viktigste forskjellen mellom skriveoppgavene i 2005 og de etter 2010 hadde med innholdsdimensjonen å gjøre: Oppgavene etterspurte innhold som elevene trolig hadde kunnskap om, for eksempel det å være elev i norsk skole, slik at prøvene kunne gjennomføres uten forberedelseshefter eller annet materiale som stilte krav til elevenes leseferdigheter.

Drøfting av oppgavene. Det er minst tre aspekter ved de oppgavene som inngår i prøvene i skriving som grunnleggende ferdigheter, som fortjener en drøfting.

For det første var oppgaveformatet ei utfordring. På eksamen i norsk og i de nasjonale prøvene i 2005 ble det utviklet integrerte oppgaver, altså oppgaver som er basert på tekster som elevene leser, og disse skal ideelt sett bidra til at alle elever har noe å skrive om, det vil si at de har tilgang til relevant

informasjon som grunnlag for å generere innhold. Dessuten ligner disse oppgavene i vesentlig grad oppgaver i skolen forøvrig; fra tidlig alder skriver elevene korte og lengre tekster basert på leste tekster. Men med integrerte oppgaver vil lesekompetansen til elevene føre til ulike muligheter for å tilegne seg materialet for å anvende det i sine egne tekster. Dermed risikerer man å måle andre ferdigheter enn skriveferdigheten, noe Messick (1996) kaller konstrukt-irrelevant varians.

Med det oppgaveformatet som ble innført i 2010, økte man sjansen for at alle elever hadde noe å skrive om, og slik forsøkte man å unngå å teste andre ferdigheter enn skriving (som for eksempel lesing). Prisen for dette valget var en type oppgave som lå langt fra skolevirkeligheten. Selv om temaet for oppgavene med enkelthet kunne knyttes til læreplaner i fag, er det mulig at oppgaveformatet i begrenset grad speilet skriving som forekom i de enkelte fagene i skolen. Dette vil i seg selv representere en trussel mot validiteten, men nå i den forstand at ekstrapoleringsmulighetene blir svakere (Kane, 2013).

På den ene siden har det vært nødvendig å standardisere prosedyrer for gjennomføring av prøvene, det vil si at prøva blir gjennomført på samme måte i alle klasserom for at elevtekstene skal være sammenlignbare, noe som er en forutsetning både for å utvikle mestringsnivåbeskrivelser og for å kunne ha pålitelig informasjon om hvordan elever presterer nasjonalt (nasjonale nivå). På den andre siden tilsier forskning innen skriveidaktikk at elevene bør lære å skrive / skrive for å lære i prosesser der de får støtte og veiledning av lærere før og underveis i skriveprosessen. Sagt på en annen måte: For at det skal være mulig for lærere å dra nytte av mestringsnivåbeskrivelsene og muligheten til å sammenligne lokale resultater mot nasjonale resultater, må skriveprøva gjennomføres med prosedyrer som ikke modellerer det som ellers kan betegnes som god skriveopplæring. Likevel ble det gitt forrang til hensynet til at alle elever gjennomførte prøva under samme forhold, og dette var også en forutsetning for at det skulle være mulig å sammenligne tekster.

Et tredje aspekt var relatert til antallet oppgaver. I tråd med konstruktdefinisjonen inngikk ikke bare flere aspekter av skriving (skriver-leser-interaksjon, innhold, tekststruktur osv.), men også flere skrivehandlinger i prøva. I 2005 skrev elevene som tok prøva to oppgaver, i perioden 2010–2015 besvarte elevene bare éi oppgave, og igjen i 2016 svarte alle elever som tok prøva på to oppgaver. I alle prøver før 2016 svarte den ene delen av utvalget på éi oppgave som var basert på éi skrivehandling, og den andre delen av utvalget på éi annen skrivehandling. Det at prøvene i et år ikke inkluderte alle skrivehandlingene, men bare to skrivehandlinger, innebar at prøva, i strikt forstand, hadde begrensede muligheter til å si noe om tilstanden i landet og om prestasjonen til den enkelte eleven. I 2016 inngikk flere skrivehandlinger, slik at det nasjonale nivået representerer bredden i norske elevers skrivekompetanse. Av praktiske grunner forble det imidlertid umulig å la den enkelte eleven svare på oppgaver tilsvarende alle skrivehandlingene i Skrivehjulet.

Fra forventningsnormer til mestringsnivåbeskrivelser

Den nasjonale prøva i skriving som grunnleggende ferdighet i 2005 skulle gjennomføres av elevenes lærere, og lærerne skulle vurdere tekstene fra egne elever. I prøvepakka inngikk veiledninger for hvordan de skulle vurdere tekstene. Veiledninga inneholdt sju vurderingsområder (figur 3): generelt førsteinntrykk, kommunikasjon, innhold, tekstopppbygging, språkbruk, rettskriving og tegnsetting, og bruk av skriftmediet (Thygesen et al., 2007). Disse områdene ble valgt for å dekke de mest sentrale dimensjonene ved elevtekster, med unntak av det første området som dreide seg om lærerens holistiske forståelse.

Vurderinga av elevtekstene skulle følge prinsippet for analytisk vurdering, noe som innebar at hver tekst skulle få seks resultat (generelt førsteinntrykk skulle ikke rapporteres videre). For hvert vurderingsområde var det tre skalanivåer, der det midterste tilsvarte «omtrent som man kan forvente av de fleste elevene på årstrinnet», og nivåene over og under tilsvarte «mye over forventet nivå» og «mye under forventet nivå». I vurderingsskjemaet (Thygesen et al., 2007, vedlegg 4) fantes fire såkalte ledespørsmål, for eksempel til området språkbruk: «Har teksten ordvalg og setningsstrukturer som passer i en faglig beskrivelse [...]?» Dessuten hadde prøvekonstruktørene utviklet «Trinnvise mål for utvikling av skriveferdighet gjennom utdanningssystemet» (Thygesen et al., 2007, s. 123ff), som var en norm som skalaen skulle forholdes til; en besvarelse som samsvarte med de beskrevne målene skulle vurderes til «omtrent som man kan forvente av de fleste elevene på årstrinnet» (se også figur 3).

Vurderingsområde	Ledespørsmål	Vurderingsskala
Førsteintrykk		
*Kommunikasjon *Innhold *Tekstopppbygging *Språkbruk *Rettskriving og tegnsetting *Bruk av skriftmediet	f.eks.: Har teksten en passende rettskriving?	<ul style="list-style-type: none"> ● Mye over det en kan forvente ● Omtrent som en kan forvente av de fleste elever på årstrinnet ● Mye under det en kan forvente
Trinnvise mål		
Eksempeltekster med kommentarer		

Figur 3. Vurderingsverktøy år 2005. Verktøyet inneholdt 6 + 1 vurderingsområder, spørsmål til respektive områder og tre nivåer. Som fundament for redskapet lå trinnvise mål, dvs. beskrivelser av hva som var rimelig å forvente av elever på årstrinnet. Videre fantes illustrasjoner på vurderingsverktøyet gjennom vurderte elevtekster med kommentarer.

Skalaen skulle altså ideelt sett generere informasjon som kunne anvendes av både læreren, skoleeier og utdanningsmyndigheter. Førstnevnte kunne få oversikt over hvor mange elever som plasserte seg på ulike sider av forventet nivå. De sistnevnte kunne få et overblikk over lokale resultater målt mot nasjonalt nivå. I tillegg ønsket man at skalaen skulle gjøre det mulig å sammenligne resultater fra år til år (Thygesen et al., 2007). I evalueringa av de nasjonale prøvene i skrivning ble det rettet kritikk mot at midtkategorien i skalaen var for stor og at det var for mange vurderingsområder (Lie et al., 2005, s. 73ff). En stor midtkategori gjorde at få lærere fikk nyansert informasjon om elevenes skrivekompetanse. Den tekniske rapporten fra prøva viste at de fleste besvarelsene ble plassert i midtkategorien: i gjennomsnitt, uavhengig av årstrinn og vurderingsområde, ble denne kategorien anvendt i 74,7 % av tilfellene. Prøvekonstruktørene var enige i at skalaen var for kort, og dessuten hadde man indikasjoner på at skalaen ikke samsvarte med læreres forståelse av typiske vurderingsskalaer i skolen. Lærerne var, i følge prøvekonstruktørene, vante til ordinalskalaer, mens den tredelte skalaen i de nasjonale prøvene kunne oppfattes som en dikotomskala, altså forventet nivå eller ikke forventet nivå (Thygesen et al., 2007, s. 90).

Etter at prøvevirksomheten gjenoppstod i 2010, ble vurderingsverktøyet betydelig endret. Prøvene skulle ha seks vurderingsområder: kommunikasjon, innhold, tekststruktur, språkbruk, rettskriving og tegnsetting, og skalaene skulle være av ulik lengde på de ulike årstrinnene. På 5. årstrinn skulle skalaen bestå av tre nivå og på 8. årstrinn fem nivå. Oppdragsbrevet tilsa også at prøvegruppa skulle utvikle mestringsnivåbeskrivelser for hvert av vurderingsområdene, i alt 18 beskrivelser for 5. årstrinn og 30 beskrivelser for 8. årstrinn. Dette innebar at skalaene var mer tydelige og konkrete enn skalaene fra 2005: for å plassere en tekst på et visst nivå skulle man, i stedet for å ta stilling til om teksten var som forventet av elever på årstrinnet, heller finne den nivåbeskrivelsen som best passet med trekk i elevteksten.

Utviklinga av mestringsnivåbeskrivelser bygde på materialet fra utvalgsprøver i skrivning. Det ble opprettet grupper med vurderingsforskere og ansatte ved Skrivesenteret som hadde ansvar for hvert sitt vurderingsområde. Gruppene vurderte og rangerte et utvalg tekster på fem nivå, og deretter ble tekstene analysert for å finne trekk i tekstene som svarte til det enkelte nivået. Denne tilnærminga innebar på den ene siden at mestringsnivåbeskrivelsene var empirisk fundert i egenskaper ved faktiske elevtekster. På den andre siden innebar denne tilnærminga en risiko for at tekstegenskaper som ikke forekom i materialet, ikke ble fanget opp og beskrevet i mestringsnivåbeskrivelsene.

Da Utdanningsdirektoratet lanserte de læringsstøttende prøvene i 2014, fungerte dessuten mestringsnivåbeskrivelsene som «nasjonale nivå», men uten at det ble publisert resultater for hvordan elevresultater ble fordelt på de ulike nivåene. Derimot beskrev det høyeste nivået innen det enkelte vurderingsområdet for respektive trinn et teoretisk maksimum for elevprestasjoner. Denne ordninga gjorde det umulig for lærere og andre interesserte å vite hvor vanlig det var at

elevtekster ble vurdert til å samsvare med det øverste nivået i skalaen. Ordninga gjorde det også umulig å la utvalgsprøvene si noe om elevers skriveutvikling over tid. To år senere ble denne ordninga omgjort, og i 2016 publiserte Utdanningsdirektoratet oversikter som viste hvordan resultater fordelte seg blant elever på femte og åttende årstrinn (Skar, 2017).

I forkant av lanseringa av de læringsstøttende skriveprøvene i 2017 ble vurderingsredskapene ytterligere justert.² Det ble utviklet en *felles* skala for femte og åttende årstrinn gjennom ei lignende tilnærming som lå til grunn for utvikling av skalaene i prøvene fra 2014 (Aasen & Skar, 2018). Den siste versjonen av vurderingsredskapet hadde fem vurderingsområder på fem nivåer og ble komplettert med ei liste med begrepsdefinisjoner (se vedlegg 3).

Drøfting av vurderingsskalaene. Vurderingsskalaer og andre vurderingsverktøy (som f.eks. kommenterte elevtekster) har ei avgjørende betydning for både begrepsvaliditet (Messick, 1996) og for hvilken informasjon ei prøve genererer. Dessuten fungerer skalaer i vurdering av skrivning som den viktigste og tydeligste operasjonaliseringa av et testkonstrukt (Knoch, 2011). Samtidig avslører vurderingsskalaer noe om forhåpningene om hva skalaen skal kunne gjøre. I 2005 var det meninga at skalaene skulle kunne anvendes for å komme med direkte og substansielle påstander om elevenes skriveferdigheter, for eksempel at 82,2 % av elevene på 4. årstrinn inkluderte relevant innhold i tekstene sine. Det var også meninga at det skulle være mulig å sammenligne resultatene fra år til år.

I 2017 var tolkninga av resultatene betydelig mindre direkte. For det første inneholdt ikke prøva forventningsnormer av samme type som i 2005, det vil si normer som uttrykte hva slags skrivekompetanse samfunnet forventet at elever utvikler i løpet av de første fire og sju årene på skolen. For det andre var det enkelte resultatet vanskeligere å tolke; selv om resultatet kunne si noe om egenskapene til en elevtekst på et visst nivå (beskrevet i mestringsnivå-beskrivelsene), var det krevende for lærere å trekke slutninger om hva dette betydde i et større perspektiv. Spørsmålet prøva ikke kunne svare på, var hvordan resultatene på prøva (f.eks. fordelinger av elever på ulike mestringsnivåer) relaterte til samfunnets forventninger om hva elever bør kunne mestre på bestemte årstrinn.

For å utnytte hele det potensialet som ligger i å ha nasjonale skriveprøver, skulle det trolig være hensiktsmessig å normere skalaen også ut ifra samfunnets forventninger til elevers skriveutvikling ved visse årstrinn etter normal skriveundervisning. At lærere og andre interessenter kjenner til hvilke nivåer som er forventet ved visse tidspunkter i skoleløpet, skulle kunne være et viktig supplement til den empiriske fordelinga av elevsvar på ulike nivå.

² Som nevnt innledningsvis stoppet Utdanningsdirektoratet utviklinga av læringsstøttende prøver i skrivning i 2016. Skrive-senteret fikk lov til å bruke data fra utvalgsprøven i 2016 til selv å utvikle og lansere læringsstøttende prøver i skrivning i 2017.

Fra konsensusrasjonale til Rasch-analyse

Den største forandringa som skriveprøvene gikk gjennom i tida mellom 2003 og 2016 var utvilsomt gjennomføringa av vurderingene. Som nevnt ovenfor, var det kvaliteten på vurderingene som medførte at skriveprøvene ble skrinlagt mellom 2006 og 2010. Ordninga med at kun én lærer vurderte elevenes tekster (og i tillegg sine egne elever) var i strid med resultater fra mange års forskning som hadde vist at vurderere sjelden kommer fram til samme resultat (f.eks. Björns-son, 1960; Edgeworth, 1890; McNamara, 1996; Weigle, 1998). Også en norsk studie (Berge, 1996) hadde påvist lav reliabilitet i vurdering av skrijving, i dette tilfellet i sensuren av norskeksamen. Beslutninga om å la elevenes lærere vurdere prøva kunne sammenfattes i en modell som ser slik ut: $X = T$, der X er observert poeng og T er det sanne poenget (med T som i «true score»). Ifølge en slik modell er det alltid perfekt samsvar mellom det resultatet en elev får på ei prøve og elevenes «sanne» resultat. På mange måter passet denne modellen også til de skalaene som ble konstruert; gjennom prøva skulle man oppnå et direkte og meget presist estimat på hvor mange elever som befant seg under, på og over forventet nivå.

I evalueringa av de nasjonale prøvene i skrijving (Lie et al., 2005) ble en annen og mer anerkjent modell brukt, nemlig modellen i den klassiske testteorien: $X = T + E$, der X og T er det samme som tidligere og E er en feilterm (med E som i «error»). I den sistnevnte ligger for eksempel reliabilitetsproblem som lav vurderersamstemmighet og lav oppgavereliasitet. Evalueringa viste at interbedømmerreliabiliteten var altfor lav, på det laveste Cohen's kappas $= 0,11$ og som høgest Cohen's kappas $= 0,37$. Konklusjonen av reliabilitetsundersøkelsen var at resultatet var upålitelig, og det innebar at resultatene på prøvene var meningsløse.

Da prøva gjenoppstod i 2010, gikk man ut fra den klassiske testteori-modellen. Målet var å redusere E -termen til et minimum gjennom delvis å la hver elevtekst bli vurdert flere ganger og delvis å etablere et såkalt tolkingsfelleskap. Begrepet «tolkingsfelleskap» ble introdusert av Fish (1980) og ble senere anvendt av White (1984) og Berge (2002) for å betegne ei gruppe med vurderere som deler oppfatning av hvordan skrijving skal vurderes og hvilke tekstegenskaper som kan relateres til ulike kompetansenivåer.

Man så for seg at to tolkingsfelleskap skulle etableres. På kort sikt skulle et fellesskap oppstå blant de 40–100 lærerne som var del av det nasjonale vurdererpanelet, altså det panelet som vurderte tekster fra utvalgsprøvene i skrijving. På lengre sikt skulle arbeidsmåten i panelet kunne overføres til alle lærere slik at alle norske lærere skulle utgjøre et tolkingsfelleskap (Skar & Jølle, 2017).

Det nasjonale vurdererpanelet ble samlet ca. to ganger i året og gikk gjennom omfattende skolering som bestod av forelesninger om vurdering av elevtekster og, framfor alt, omfattende parvurdering (se f.eks. Jølle, 2014, 2015 for inngående redegjørelser). Parvurdering, som hadde mye til felles med eksamenssensur, innebar at hver vurderer først vurderte et visst antall tekster for så å

diskutere og bli enige om vurderinga sammen med en vurderingspartner. Sammensetninga av vurdererpar ble utført gjennom et rotasjonsprinsipp for at hver vurderer skulle eksponeres for så mange vurderingspartnere som mulig. Hver tekst ble lest av fire vurderere, men ettersom det var vurdererparets oppgave å bli enige om et resultat, fikk hver tekst bare to resultat. Det var disse to resultatene som ble anvendt for å estimere reliabiliteten. Den var ikke tilfredsstillende. I 2010 ble reliabiliteten målt til $ICC = 0,46$ (Fasting, 2011) og i 2014, det siste året med samme modell, til $ICC = 0,59$ (Skar et al., 2015).

På våren 2014 ble det gjennomført ei ny utprøving, men denne gangen ble det innført en ny statistisk modell, nemlig den mangefasetterte Rasch-modellen (Barkaoui, 2014; Linacre, 2017; Rasch, 1980). I følge denne modellen var elevens resultat avhengig av vanskelighetsgraden til oppgava og vurderernes strenghet. Denne modellen åpnet for at elevresultatene ble justert for de faktorene som inngikk i modellen, noe som innebærer at en elevtekst som ble vurdert av en streng vurderer, kunne få et høyere skalapoeng enn råpoeng. Slik sett representerte modellen et paradigmeskifte; i stedet for å minimere feiltermen (E) skulle man estimere den og aktivt anvende den for å justere elevresultat. Modellen innebar også at man gikk vekk fra idealet om et nasjonalt tolkingsfellesskap.

Drøfting av vurderingene. Da prøvevirksomheten ble avsluttet i desember 2016, hadde man redusert usikkerheten rundt resultatenes reliabilitet. Samtidig hadde man gått fra en tanke om at et tolkingsfellesskap var til stede, eller var mulig å styrke gjennom skolering og samvurdering, til en vesentlig annerledes forståelse av hva tolkingsfellesskap blant norske lærere kan være. I et slikt fellesskap er ikke lærere utbyttbare med hverandre, men kompletterer hverandre, det vil si at ulike læreres kompetanse og normer brytes mot hverandre. Og hver gyldige vurdering vil kreve flere uavhengige lesere. Gyldig vurdering krever også estimering av hvordan lærere bruker vurderingsverktøyene forskjellig, og endelig justering som tar høyde for dette. De tradisjonelle verktøyene (som f.eks. kriterier og eksempeltekster) bør derfor suppleres med statistiske verktøy i vurdering der mye står på spill.

Avslutning

Formålet med denne artikkelen har vært å løfte fram og samle noen av de sentrale innsiktene fra arbeidet med utvikling av nasjonale og læringsstøttende prøver i skrijving. Vi har gjort det gjennom å presentere og drøfte sentrale utviklingstrekk i arbeidet med prøvene.

Artikkelen har pekt på fortsatte utfordringer med å ha oppgavesett som fanger bredden og dybden i konstruktdefinisjonen og som gjenspeiler god skriveidaktikk. Dessuten har vi vist at de nye skalaene for så vidt kunne brukes for å vurdere elevenes skrivekompetanse, men at meninga i et gitt testresultat

ikke var helt klar på grunn av manglende kopling til samfunnets forventinger (dvs. offisielle normer for skriveutvikling gjennom hele skoleløpet). I artikkelen har vi videre vist at prøveutviklerne lyktes med å definere et testkonstrukt som fanget bredden i skrivning, og vi har også vist at prøveutviklerne lyktes med å løse problemet med lavt samsvar mellom de som vurderte tekstene ved å innføre en Rasch-analytisk tilnærming.

Innledningsvis snakket vi om at skrivning gjerne kan bli omtalt som problem-barnet i vurderingsklassa; det er vanskelig å måle elevers skrivekompetanse på en gyldig måte. Selv om vi har gjort rede for noen utfordringer ved utvikling av skriveprøver, har flere av problemene i skriveprøvenes tidlige fase blitt løst. Alternativene til skriveprøver vil være enten å måle skrivekompetanse indirekte gjennom for eksempel grammatikkprøver, noe som vil føre til en snever forståelse av hva skrivning er, eller å ikke skaffe informasjon om elevers skrivekompetanse i det hele tatt, noe som vil føre til at skrivning risikerer å bli nedprioritert i skolen. En slik forsnevring og nedprioritering bryter med anerkjennelsen av at skrivning er et sentralt verktøy for læring og for vurdering av elevenes kompetanse i fagene i skolen. Med den kompetansen som fagmiljøer i Norge har i dag, og som denne artikkelen viser, vil det være mulig å utvikle bedre skriveprøver i framtida enn hva man kunne for bare 15 år siden.

Om forfatterne

Gustaf Bernhard Uno Skar er førsteamanuensis i norsksdidaktikk ved Norges teknisk-naturvitenskapelige universitet. Hans forskningsinteresser omfatter blant annet vurdering og skrivevurdering.

Institusjonstilknytning: Nasjonalt senter for skriveopplæring og skriveforskning, Norges teknisk-naturvitenskapelige universitet, NTNU ILU, 7491 Trondheim.

E-post: gustaf.b.skar@ntnu.no

Arne Johannes Aasen er senterleder ved Nasjonalt senter for skriveopplæring og skriveforskning. Hans forskningsinteresser omfatter blant annet skriveidaktikk og vurdering.

Institusjonstilknytning: Nasjonalt senter for skriveopplæring og skriveforskning, Norges teknisk-naturvitenskapelige universitet, NTNU ILU, 7491 Trondheim.

E-post: arne.j.aasen@ntnu.no

Referanser

Anderson, S. (2015). *Lang dags ferd fra tradisjon til fornyelse*. Hentet 15.06.2018 fra http://udirbeta.udir.no/wp-content/uploads/2015/08/Lang-dags-ferd_-_Sissel_Anderson_blogg-00000002.pdf

- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (2010). *Language assessment in practice: developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Barkaoui, K. (2014). Multifaceted Rasch analysis for test evaluation. I A. J. Kunnan (red.), *The Companion to Language Assessment* (s. 1301–1322). Chichester, West Sussex: Wiley-Blackwell.
- Bazerman, C., Applebee, A. N., Berninger, V. W., Brandt, D., Graham, S., Matsuda, P. K., ... Schleppegrell, M. (2017). Taking the Long View on Writing Development. *Research in the Teaching of English*, 51(3), 351–360.
- Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Berge, K. L. (1996). *Norsksensorenes tekstnormer og doxa: en kultursemiotisk og sosiotekstologisk analyse*. Doktoravhandling, Norges teknisk-naturvitenskapelige universitet, Trondheim.
- Berge, K. L. (2002). Hidden Norms in Assessment of Students' Exam Essays in Norwegian Upper Secondary Schools. *Written Communication*, 19(4), 458–492. doi: <http://doi.org/10.1177/074108802238011>
- Berge, K. L. (2005a). Skrivning som grunnleggende ferdighet og som nasjonal prøve – ideologi og strategier. I A. J. Aasen & S. Nome (red.), *Det nye norskfaget* (s. 161–188). Bergen: Fagbokforlaget.
- Berge, K. L. (2005b). Studie 3: Skriveprøvens pålitelighet. I K. L. Berge, L. S. Evensen, F. Hertzberg & W. Vagle (red.), *Ungdommers skrivekompetanse. Bind I: Norsksensuren som kvalitetsvurdering* (s. 101–113). Oslo: Universitetsforlaget.
- Berge, K. L. (2007). Appendiks. Hvorfor ble de nasjonale prøvene i skrivning et mislykket prosjekt? Historien om et kvalitetssikringsnotat om de nasjonale skrivprøvene rettet til statsråd Kristin Clemet. I S. Tveit (red.), *Elevvurdering i skolen: grunnlag for kulturendring* (s. 245–248). Oslo: Universitetsforlaget.
- Berge, K. L. (2015). *Skrivevurdering – tradisjon og fornyelse*. Paper presented at the First Nordic Symposium on Testing and Assessing Writing (NTAW), Trondheim, Norway, September 20–29, 2015.
- 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. doi: <http://doi.org/10.1080/09585176.2015.1129980>
- Björnsson, C. H. (1960). *Uppsatsbedömning och uppsatsskrivning*. Stockholm: Almqvist & Wiksell.
- 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. doi: <http://doi.org/10.1177/0265532214542994>
- Breland, H. M., Bridgeman, B. & Fowles, M. E. (1999). *Writing Assessment in Admission to Higher Education: Review and Framework*. New York: College Entrance Examination Board.
- Camp, H. (2012). The psychology of writing development—And its implications for assessment. *Assessing Writing*, 17(2), 92–105. doi: <http://doi.org/10.1016/j.asw.2012.01.002>
- Dagsland, S. (2018). *Om å jakte på heffalomper: Et metalingvistisk perspektiv på antatt reflekterende og antatt utforskende skrivning i norsk og matematikk*. Doktoravhandling, Norges teknisk-naturvitenskapelige universitet, Trondheim. Hentet fra https://brage.bibsys.no/xmlui/bitstream/handle/11250/2494191/SindreDagsland_PhD.pdf?sequence=1&isAllowed=y

- Edgeworth, F. Y. (1890). The Element of Chance in Competitive Examinations. *Journal of the Royal Statistical Society*, 53(3), 460–475, 644–663.
- Evensen, L. S. (2010). En gyldig vurdering av elevers skrivekompetanse? I J. Smidt (red.), *Rammer for skriving: om skriveutvikling i skole og yrkesliv* (s. 13–31). Trondheim: Tapir akademisk forlag.
- Fasting, R. B. (2011). *Nasjonale utvalgsprøver i skriving som grunnleggende ferdighet. Vurdererfellesskap og reliabilitet ved vurdering av skriving*. Upublisert. Trondheim: Nasjonalt senter for skriveopplæring og skriveforskning.
- Fasting, R. B., Thygesen, R., Berge, K. L., Evensen, L. S. & Vagle, W. (2009). National Assessment of Writing Proficiency Among Norwegian Students in Compulsory Schools. *Scandinavian Journal of Educational Research*, 53(6), 617–637.
doi: <http://doi.org/10.1080/00313830903302117>
- Fish, S. (1980). *Is There a Text in This Class? The Authority of Interpretive Communities*. Cambridge, Mass.: Harvard University Press.
- Flower, L. & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365–387. Hentet fra <http://www.jstor.org/stable/356600>
- Huot, B. & Perry, J. (2009). Toward a New Understanding for Classroom Writing Assessment. I R. Beard (red.), *The Sage handbook of writing development* (s. 423–435). Los Angeles: SAGE.
- Jeffrey, J. V., Elf, N., Skar, G. B. & Wilcox, K. (u.u.). Writing Development and Education Standards in Cross-national Perspective. *Writing & Pedagogy*.
- Jølle, L. (2014). Pair assessment of pupil writing: A dialogic approach for studying the development of rater competence. *Assessing Writing*, 20, 37–52.
doi: <http://doi.org/10.1016/j.asw.2014.01.002>
- Jølle, L. (2015). *Vurderingsdialogen. En undersøkelse av tekstvurderingspraksis ved nasjonal læringsstøttende prøve i skriving*. Doktoravhandling, Norges teknisk-naturvitenskapelige universitet, Trondheim. Hentet fra https://brage.bibsys.no/xmlui/bitstream/handle/11250/298034/Lennart_Jolle.pdf?sequence=3
- Jølle, L. & Skar, G. B. (u.u.). “Digging for Gold” or “Sticking to the Criteria”: Teachers’ Rationales When Serving as Professional Raters. *Scandinavian Journal of Educational Research*.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73. doi: <http://doi.org/10.1111/jedm.12000>
- Kane, M., Crooks, T. & Cohen, A. (1999). Validating Measures of Performance. *Educational Measurement: Issues and Practice*, 18(2), 5–17. doi: <http://doi.org/10.1111/j.1745-3992.1999.tb00010.x>
- Kellogg, R. T. (2008). Training writing skills: A cognitive development perspective. *Journal of Writing Research*, 1(1), 1–26. doi: <http://doi.org/10.17239/jowr-2008.01.01.1>
- Knoch, U. (2009). *Diagnostic Writing Assessment: The Development and Validation of a Rating Scale*. Frankfurt am Main: Peter Lang.
- Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, 16(2), 81–96.
doi: <http://doi.org/10.1016/j.asw.2011.02.003>
- Krogh, L. C. (2016). *Kreativitet og ambivalens : en undersøkelse av variasjoner i vurdering og kjennetegn ved sprikvurderte tekster fra eksamen i norsk hovedmål 2015*. Masteroppgave, Høgskolen i Sørøst-Norge. Hentet fra <http://hdl.handle.net/11250/2412785>

- Lie, S., Hopfenbeck, T., Ibsen, E. & Turmo, A. (2005). *Nasjonale prøver på ny prøve*. Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.
- Linacre, J. M. (2017). *A user's guide to FACETS. Rasch-model computer programs. Program manual 3.80.0*. Hentet 25.05.2017 fra <http://www.winsteps.com/a/Facets-ManualPDF.zip>
- McNamara, T. F. (1996). *Measuring second language performance*. New York: Longman.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. doi: <http://doi.org/10.1177/026553229601300302>
- Myford, C. M. & Wolfe, E. W. (2003). Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part I. *Journal of Applied Measurement*, 4(4), 386–422.
- Norendal, A. (2016). Dannelsingsperspektiver i grunnskolens eksamensoppgaver i norsk. *Acta Didactica Norge*, 10(3), Art. 4.
- NOU 2002:10 (2002). *Førsteklasses fra første klasse – Forslag til rammeverk for et nasjonalt kvalitetsvurderingssystem av norsk grunnopplæring*. Oslo: Utdannings- og forskningsdepartementet. Hentet 09.12.2018 fra <https://www.regjeringen.no/no/dokumenter/nou-2002-10/id145378/>
- Otnes, H. (2015). Skriveoppgaver under lupen. I H. Otnes (red.), *Å invitere elever til skriving* (s. 11–27). Bergen: Fagbokforlaget.
- Purpurra, J. E. (2010). Assessing communicative language ability: models and their components. I E. Shohamy & N. H. Hornberger (red.), *Encyclopedia of language and education. Vol. 7, Language testing and assessment* (s. 53–68). New York: Springer.
- Purves, A. C., Gorman, T. P. & Takala, S. (1988). The Development of the Scoring Scheme and Scales. I T. P. Gorman, A. C. Purves & R. E. Degenhart (red.), *The IEA Study of Written Composition I: The International Writing Tasks and Scoring Scales* (s. 41–58). Oxford: Pergamon.
- Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: The University of Chicago Press.
- Rijlaarsdam, G., Van den Bergh, H., Couzijn, M., Janssen, T., Braaksma, M., Tillema, M., Van Steendam, E. & Raedts, M. (2012). Writing. I K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major & H. L. Swanson (red.), *APA educational psychology handbook, Vol 3: Application to learning and teaching* (s. 189–227). Washington: American Psychological Association. doi: <http://doi.org/10.1037/13275-009>
- Rose, D. (2016). New Developments in Genre-Based Literacy Pedagogy. I C. A. MacArthur, S. Graham & J. Fitzgerald (red.), *Handbook of writing research* (2nd ed.) (s. 227–242). New York & London: The Guilford Press.
- Shanahan, T. & Shanahan, C. (2008). Teaching Disciplinary Literacy to Adolescents: Rethinking Content-Area Literacy. *Harvard Educational Review*, 78(1), 40–59. doi: <http://doi.org/10.17763/haer.78.1.v62444321p602101>
- Skar, G. B. (2017). *The Norwegian National Sample-Based Writing Test 2016: Technical Report*. Trondheim: Nasjonalt senter for skriveopplæring og skriveforskning. Hentet fra <http://www.skrivesenteret.no/uploads/files/Skriveproven2017/NSBWT2017.pdf>
- Skar, G. B., Evensen, L. S. & Iversen, J. M. (2015). *Læringsstøttende prøver i skriving 2014. Teknisk rapport*. Trondheim: Nasjonalt senter for skriveopplæring og skriveforskning.
- Skar, G. B. & Iversen, J. M. (2016). *Teknisk rapport: Pilotering av oppgaver til den nasjonale utvalgsprøven i skriving 2016*. Trondheim: Nasjonalt senter for skriveopplæring og skriveforskning.
- Skar, G. B. & Jølle, L. (2017). Teachers as raters: Investigation of a long term writing assessment program. *L1 Educational Studies in Language and Literature*, 17(Open Issue), 1–30. doi: <http://doi.org/10.17239/L1ESLL-2017.17.01.06>
- Skar, G. B. U. & Aasen, A. J. (2016). Risikotrekk og skjulte kvalitetar i elevtekstar. *Nordic Journal of Literacy Research*, 2(1), 1–20. doi: <http://doi.org/10.17585/njlr.v2.199>

- Skjelten, S. M. (2013). *Jakta på kvalitetsforskjellar i elevane sine tekstar. Kva skil gode tekstar frå middels gode?* Doktoravhandling, Universitetet i Oslo.
- Solheim, R. & Matre, S. (2014). Forventninger om skrivekompetanse. Perspektiver på skriving, skriveopplæring og vurdering i «Normprosjektet». *Viden Om Læsning*, 8(15), 76–88.
- St.meld. nr. 30 (2003–2004). *Kultur for læring*. Oslo:Utdannings- og forskningsdepartementet. Hentet 10.12.2018 fra <https://www.regjeringen.no/contentassets/988cdb018ac24eb0a0cf95943e6cdb61/no/pdfs/stm200320040030000dddpdfs.pdf>
- St.prp nr. 1, Tillegg nr. 3 (2002–2003). Oslo: Kunnskapsdepartementet. <https://www.regjeringen.no/no/dokumenter/stprp-nr-1-tillegg-nr-3-2002-2003-/id435850/>
- Thygesen, R., Evensen, L. S., Berge, K. L., Fasting, R. B., Vagle, W. & Haanæs, I. R. (2007). *Nasjonale prøver i skriving som grunnleggende ferdighet. Sluttrapport*. Nasjonalt senter for leseopplæring og leseforskning, Universitetet i Stavanger.
- Utdanningsdirektoratet (2014). *Signert avtale om utvikling av læringsstøttende prøver i skriving, Bilag 4: Rammeverk for skriveprøver*, 28.05.2014.
- Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2), 263–287. doi: <http://doi.org/10.1177/026553229801500205>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- White, E. M. (1984). Holisticism. *College Composition and Communication*, 35(4), 400–409. Hentet fra <http://www.jstor.org/stable/357792>
- Aasen, A. J. & Skar, G. B. (2018). *Developing Scales and Investigating Writing Proficiency Among Fifth- and Eighth-Grade Students in Norway*. Presentert på Annual Meeting of the American Educational Research Association, New York City, 13–17.04.2018.

Vedlegg 1

Oppgaver for elever på 10. trinn i nasjonale prøver i skrijving som grunnleggende ferdighet 2005

Oppgave 1 (utforsking)

Utforskingen av planeten Mars reiser på nytt spørsmål om det kan ha vært liv på planeten, og om det fortsatt kan være liv der.

Skriv en fagtekst der du gjør greie for hva som taler for, og hva som taler imot at det er eller kan ha vært liv på planeten. Gi din vurdering av hva som virker mest sannsynlig. Lag en tittel til teksten din.

Oppgave 2 (refleksjon)

Til alle tider har stjernene, planetene og verdensrommet inspirert mennesker til å tenke omkring "de store spørsmålene" og det å være menneske:

Hvem er jeg, og hvorfor er jeg her? Hva er meningen med livet? Hva kan vi vite sikkert egentlig? Hva er rett, og hva er galt? Hva er vakkert, og hva er stygt? Moder jord – hva er det vi gjør med vår blå-grønne planet?

Dette er bare noen av de evige spørsmål som himmelhvelvingen kan få en til å fundere over. Kanskje har du fundert over slike spørsmål en gang iblant – kanskje i forlengelsen av tanker og ideer du er blitt kjent med gjennom KRL-faget?

Kanskje har du egne erfaringer som gjør at disse eller andre spørsmål er blitt viktige for deg? Skriv en reflekterende tekst der du deler med andre mennesker dine tanker omkring ett av "de store spørsmålene". Gi teksten din en tittel.

Oppgavene er hentet fra Thygesen et al. (2007, s. 78).

Vedlegg 2

Eksempel på oppgave i utvalgsprøva i skrijving 2015

Ein barndomsvenn av deg har gått på skule i utlandet heilt sidan 1. klasse. Snart skal denne vennen flytte attende til Noreg og begynne på skulen din. Ho eller han har spurt deg korleis skulekvardagen på skulen dykkar er. Du svarar med å beskrive skulekvardagen for han eller ho.

Ein venn kjem heim frå utlandet og skal begynne på skulen din. Beskriv skulekvardagen for vennen din.

Se Skar og Aasen (2016, s. 8f) for mer informasjon.

Ein barndomsvenn av deg har gått på skule i utlandet heilt sidan 1. klasse.

Snart skal denne vennen flytte attende til Noreg og begynne på skulen din. Ho eller han har spurt deg korleis skulekvardagen på skulen dykkar er. Du svarar med å beskrive skulekvardagen for han eller ho. Ein venn kjem heim frå utlandet og skal begynne på skulen din. Beskriv skulekvardagen for vennen din.

(Se Aasen & Skar 2016 for mer informasjon om nettopp denne oppgava.)

Vedlegg 3

Mestringsnivåbeskrivelser 2016

(Hentet fra Skar, 2017.)

Skriver-leser-interaksjon

NIVÅ 1	NIVÅ 2	NIVÅ 3	NIVÅ 4	NIVÅ 5
<ul style="list-style-type: none"> • Teksten etablerer en uklart relasjon mellom skriver og leser. For å kunne forstå teksten kreves tilgang til oppgaveformuleringen. 	<ul style="list-style-type: none"> • Den sakprøgede teksten gjøres et forsøk på å etablere en relevant relasjon mellom skriver og leser (jf. oppgaveformuleringen), men det kan være noe uklart hvorfor skriveren henverder seg til denne leseren. • Den sakprøgede teksten kan i enkelte tilfeller gjøre forsøk på å ivareta leserens behov for kjennskap til deltaker, begrep og omstendigheter (f.eks. gjennom forklaringer, som inntil nå ikke er tilpasset leseren i oppgaveformuleringen). 	<ul style="list-style-type: none"> • Den sakprøgede teksten antyder en relevant relasjon mellom skriver og leser (jf. oppgaveformuleringen), men forholder seg ikke konsekvent til denne leserens perspektiv. • Den sakprøgede teksten ivaretar i noen grad leserens behov for kjennskap til deltaker, begrep og omstendigheter (f.eks. gjennom forklaringer, som i noen tilfeller er tilpasset leseren i oppgaveformuleringen). 	<ul style="list-style-type: none"> • Den sakprøgede teksten etablerer en stort sett relevant relasjon mellom skriver og leser (jf. oppgaveformuleringen), men forholder seg ikke alltid til denne leserens perspektiv. • Den sakprøgede teksten ivaretar stort sett leserens behov for kjennskap til deltaker, begrep og omstendigheter (f.eks. gjennom forklaringer, som ofte er tilpasset leseren i oppgaveformuleringen). 	<ul style="list-style-type: none"> • Den sakprøgede teksten etablerer en relevant relasjon mellom skriver og leser (jf. oppgaveformuleringen) og forholder seg konsekvent til denne leserens perspektiv. • Den sakprøgede teksten ivaretar gjennomgående leserens behov for kjennskap til deltaker, begrep og omstendigheter (f.eks. gjennom forklaringer som er tilpasset leseren i oppgaveformuleringen).
	<ul style="list-style-type: none"> • Fiksjonsteksten iscenesetter en like sammenhengende og lite troverdig fiksjonsverden. 	<ul style="list-style-type: none"> • Fiksjonsteksten iscenesetter en fiksjonsverden med deltaker, omstendigheter og atmosfære, men som ikke er gjennomgående troverdig og som kan inneholde en del logiske brister. 	<ul style="list-style-type: none"> • Fiksjonsteksten iscenesetter en stort sett troverdig fiksjonsverden med deltaker, omstendigheter og atmosfære. 	<ul style="list-style-type: none"> • Fiksjonsteksten iscenesetter en gjennomgående troverdig fiksjonsverden med deltaker, omstendigheter og atmosfære.
		<ul style="list-style-type: none"> • Den sakprøgede teksten kan i noen få tilfeller inkludere skilting, men denne er ikke alltid nødvendig eller funksjonell. 	<ul style="list-style-type: none"> • Den sakprøgede teksten kan inkludere skilting og denne er for det meste funksjonell og bidrar til å holde leseren fokusert. 	<ul style="list-style-type: none"> • Den sakprøgede teksten kan inkludere skilting, som er brukt på en balansert og systematisk måte.

Innhold

NIVÅ 1	NIVÅ 2	NIVÅ 3	NIVÅ 4	NIVÅ 5
<ul style="list-style-type: none"> • Teksten har et innhold som er delvis relevant eller som har deler som er relevante for oppgaven. • Emnet er i fokus i deler av teksten. • Den sakpregede teksten inneholder noen få og enkle eksempler, utdypninger og begrunnelser, men de kan være lite relevante. 	<ul style="list-style-type: none"> • Teksten har et innhold som for det meste er relevant for oppgaven. • Emnet er stort sett i fokus i hele teksten. • Den sakpregede teksten inneholder enkle eksempler, utdypninger og begrunnelser, og de fleste er relevante. 	<ul style="list-style-type: none"> • Teksten har et innhold som er relevant for oppgaven. • Den sakpregede teksten inneholder noen relevante eksempler, utdypninger og/eller begrunnelser. 	<ul style="list-style-type: none"> • Emnet er i fokus i hele teksten. Innholdet er vektet på en formålstjenlig måte. • Den sakpregede teksten inneholder flere gode og relevante eksempler, utdypninger og/eller begrunnelser. 	<ul style="list-style-type: none"> • Emnet er i fokus og kan utvikles i teksten. Innholdet er vektet på en formålstjenlig måte. • Den sakpregede teksten inneholder gjennomgående gode og relevante eksempler, utdypninger og/eller begrunnelser.
<ul style="list-style-type: none"> • Fiksjonsteksten er handlingsdeven eller består av løseverne hendelser. 	<ul style="list-style-type: none"> • Fiksjonsteksten er primært handlingsdeven og kan vise tendenser til utvikling av motiv, tematikk, karakterer, miljøer eller hendelser. 	<p>••••→</p>	<ul style="list-style-type: none"> • Fiksjonsteksten kan være delvis handlingsdeven, men inneholder også deler der motiv, tematikk, karakterer, miljøer eller hendelser er godt utviklet. 	<ul style="list-style-type: none"> • Fiksjonsteksten viser utvikling av motiv, tematikk, karakterer, miljøer eller hendelser, og disse aspektene ved teksten henger godt sammen.

••••→ = samme som forrige nivå

Tekstoppbygging

NIVA 1	NIVA 2	NIVA 3	NIVA 4	NIVA 5
<ul style="list-style-type: none"> • På overordnet nivå har den sakpregede teksten en klar struktur. • Fiksjonsteksten har en klar struktur. 	<ul style="list-style-type: none"> • Den sakpregede teksten kan vise forsøk på struktur (f.eks. gjennom innledning og/eller avslutning, punkter eller mellomoverskrifter). • Fiksjonsteksten kan vise forsøk på struktur. 	<ul style="list-style-type: none"> • Den sakpregede teksten viser forsøk på bruk av strukturingsprinsipper som er tilpasset skrive-situasjonen (f.eks. oppbygging mot et hovedpoeng). • Fiksjonsteksten kan vise forsøk på bruk av strukturingsprinsipper for å utvikle plottet. 	<ul style="list-style-type: none"> • Den sakpregede teksten viser bruk av strukturingsprinsipper som gjør at den stort sett har en hensiktsmessig oppbygging tilpasset skrive-situasjonen. Teksten kan ha en avslutning (f.eks. gjennom en oppsummering eller en avsluttende appell). • Fiksjonsteksten anvender strukturingsprinsipper for å utvikle plottet. 	<ul style="list-style-type: none"> • Den sakpregede teksten viser bruk av strukturingsprinsipper som gjør at den har en hensiktsmessig oppbygging tilpasset skrive-situasjonen. Tekstens innledende del danner en relevant bakgrunn for teksten. Teksten har en hensiktsmessig avslutning (f.eks. gjennom en oppsummering eller en avsluttende appell). • Fiksjonsteksten anvender systematisk avanserte strukturingsprinsipper for å utvikle plottet.
<ul style="list-style-type: none"> • Avsnittene er korte og består ofte av enkeltstående utsagn. 	<ul style="list-style-type: none"> • Avsnittene består av innholdselementer som er stilt opp assosiativt eller uten logisk rekkefølge. Avsnittene kan bestå av enkeltstående utsagn. 	<ul style="list-style-type: none"> • Avsnittene består av innholdselementer som ofte er tematiske grupper. 	<ul style="list-style-type: none"> • Avsnittene har som regel lindre logisk struktur (f.eks. med tematiseringen) og er ofte tematiske grupper. Dette gjør teksten lett å følge (f.eks. kan forklaringer ofte presenteres i en funksjonell rekkefølge). Avsnitt kan være korrekt grafisk markert. 	<ul style="list-style-type: none"> • Avsnittene har lindre logisk struktur (f.eks. med tematiseringen). Avsnittene er tematiske ordnet på en måte som gjør at teksten blir lett å følge (f.eks. kan forklaringer ofte presenteres i en funksjonell rekkefølge – f.eks. det generelle før det spesifikke eller avslutningen kan peke tilbake på innledningen). Avsnitt er som oftest korrekt grafisk markert.
<ul style="list-style-type: none"> • Teksten kan vise bruk av flere typer enkle setningsforbindere, men disse er ikke alltid brukt på en funksjonell måte. 	<ul style="list-style-type: none"> • Teksten viser bruk av flere typer enkle setningsforbindere. Bruken er lite variert. 	<ul style="list-style-type: none"> • Teksten viser bruk av flere typer setningsforbindere. Ofte overforbrukes enkelte setningsforbindere. I noen tilfeller er setningsforbindere overløydige. 	<ul style="list-style-type: none"> • Teksten viser bruk av flere typer setningsforbindere, som for det meste er brukt på en funksjonell måte. 	<ul style="list-style-type: none"> • Teksten viser funksjonell bruk av varierte og avanserte setningsforbindere. Disse er ikke brukt overløydig.

Språkkbruk

NIVÅ 1	NIVÅ 2	NIVÅ 3	NIVÅ 4	NIVÅ 5
<ul style="list-style-type: none"> • Setningene begynner stort sett på samme måte. Teksten har enkel setningsstruktur uten underordning. 	<ul style="list-style-type: none"> • Setningene har noe variasjon i forfattet, men teksten viser likevel variert setningsstruktur, for eksempel ved mange repeterende setningsledd. 	<ul style="list-style-type: none"> • Setningene har noe variasjon i forfattet. Teksten kan vise begynnende kompleks setningsstruktur. Enkeltle setninger kan ha brister i syntaks når setningene blir komplekse. 	<ul style="list-style-type: none"> • Setningene har noe variasjon i forfattet. Delet av teksten kan vise kompleksitet og variasjon i setningsstruktur. 	<ul style="list-style-type: none"> • Setningene har variasjon i forfattet. Teksten har kompleks og variert setningsstruktur.
<ul style="list-style-type: none"> • Teksten er preget av muntlig språk med assosiativ innholdselementer, eller muntlige formuleringer. 	<ul style="list-style-type: none"> • Teksten kan være preget av muntlig språk med assosiativ innholdselementer, innskudd og mange fylford. • Det er lite variasjon i ordvalg. Begreper og formuleringer er ofte upresise. 	<ul style="list-style-type: none"> • Teksten kan vise noe variert ordvalg. Begreper og formuleringer kan være presise. 	<ul style="list-style-type: none"> • Teksten viser stort sett variert ordvalg. Det er stort sett presisjon i formuleringer og begreper. 	<ul style="list-style-type: none"> • Teksten viser variert ordvalg og presisjon i formuleringer og begreper. • I tekster der det er relevant, viser teksten riktig bruk av avanserte begreper.
←→ = samme som forrige nivå				
←→				

Formverk, rettskriving og tegnsetting

NIVÅ 1	NIVÅ 2	NIVÅ 3	NIVÅ 4	NIVÅ 5
<ul style="list-style-type: none"> • Teksten viser i stor grad bruk av fonologisk strateg. Noen ikke-lydrette ord staves riktig. 	<ul style="list-style-type: none"> • Teksten har riktig skrivemåte for noen lange lydrette ord og flere ikke-lydrette ord. Teksten kan ha noen dialektord. 	<ul style="list-style-type: none"> • Teksten har stort sett riktig skrivemåte for de fleste ord. Teksten inneholder ofte og-/å-feil, som en konsekvens av at den er mer språklig avansert. 	<ul style="list-style-type: none"> • Teksten viser riktig skrivemåte for de fleste ord. Enkelt og-/å-feil kan forekomme. 	<ul style="list-style-type: none"> • Teksten viser riktig skrivemåte for de fleste ord.
<ul style="list-style-type: none"> • Teksten inneholder noen punktum etter helsetninger. Teksten har for det meste stor bokstav i egenavn og i ny setning etter punktum. 	<ul style="list-style-type: none"> • Andre skilletegn prøves ut i tillegg til punktum: spørretegn, utropstejn. Teksten har ofte komma ved opprnsing. Teksten har for det meste stor bokstav i egenavn og i ny setning etter punktum. 	<ul style="list-style-type: none"> • Teksten viser stort sett riktig bruk av store skilletegn. Teksten har komma ved opprnsing. Teksten kan ha komma mellom helsetninger som bindes sammen med setningsforbindere. Teksten har stor bokstav i egenavn og i ny setning etter punktum. 	<ul style="list-style-type: none"> • Teksten viser stort sett riktig bruk av store skilletegn. Teksten har komma ved opprnsing og mellom de fleste helsetninger som bindes sammen med setningsforbindere. Teksten har stor bokstav i egenavn og i ny setning etter punktum. 	<ul style="list-style-type: none"> • Teksten viser riktig bruk av store skilletegn. Teksten viser stort sett korrekt bruk av komma, men feil i mer komplekse setningsstrukturer kan forekomme (f.eks. etter en som-setning). Teksten har stor bokstav i egenavn og i ny setning etter punktum.
<ul style="list-style-type: none"> • Bruk av tempus kan være lite konsekvent. 	<ul style="list-style-type: none"> • Teksten viser for det meste funksjonell bruk av tempus i enkle setninger. 	<ul style="list-style-type: none"> • Teksten viser for det meste funksjonell bruk av tempus. Enkelt feil kan forekomme i avanserte setninger. 	<ul style="list-style-type: none"> • Teksten viser korrekt samsvarende svarbøying. 	<ul style="list-style-type: none"> • Teksten viser funksjonell bruk av tempus.

→ = samme som forrige nivå